Inhaltsverzeichnis

7 Differentialrechnung mehrerer Variablen

Der Graph einer glatten Funktion $f : \mathbb{R}^2 \to \mathbb{R}$ ist eine Fläche $\mathcal{F}$ in $\mathbb{R}^3$. In der Nähe des Punktes $P \in \mathcal{F}$ kann man die Fläche durch die Tangentialebene $\mathcal{T}$ im Punkt P approximieren. Glatt bedeutet hierbei differenzierbar in einem zu präzisierenden Sinn. Wir untersuchen dazu zunächst das Änderungsverhalten von Kurven γ, die in der Fläche $\mathcal{F}$ durch einen Punkt P in Richtung der Koordinatenachsen verlaufen. Das führt zum Begriff der partiellen Ableitungen. Verlaufen die Kurven in $\mathcal{F}$ oberhalb von Geraden durch P in Richtung $v \in \mathbb{R}^2$, kommt man analog zum Begriff der Richtungsableitung in Richtung v. Um zu sehen, wie man sinnvollerweise die Differenzierbarkeit für Funktionen in Banachräumen definieren sollte, untersuchen wir zunächst als einfachsten Fall die partiellen Ableitungen von Funktionen $f : \mathbb{R}^2 \to \mathbb{R}$.

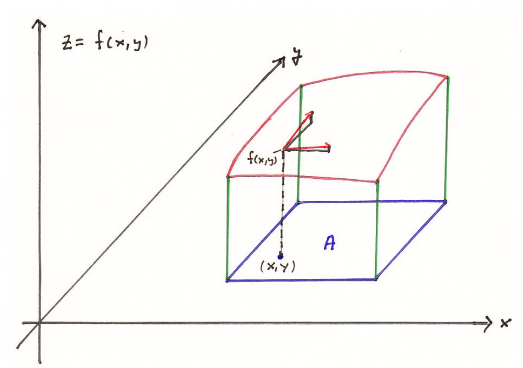
7.1 Partielle Ableitungen

Sei $f : A \subseteq \mathbb{R}^2 \to \mathbb{R}$ stetig. f repräsentiert über seinen Graphen
$\Gamma := \{(x_1, x_2, f(x_1, x_2)) \mid (x_1, x_2) \in A\}$ eine "Fläche" im $\mathbb{R}^3$.

Ist das Änderungsverhalten von f in (x_1, x_2) beschreibbar durch partielle Ableitungen, d.h. durch das Änderungsverhalten entlang Tangenten in achsenparallelen Richtungen? Sei $x = (x_1, \dots, x_n) \in \mathbb{R}^n$.

Definition. Sei $A \subseteq \mathbb{R}^n$ offen und $x \in A$. Eine Abbildung $f : A \to \mathbb{R}$ ist **in x partiell nach x_j differenzierbar** $(j \in \{1, \ldots, n\}) :\Leftrightarrow \frac{\partial f}{\partial x_j}(x) := \lim\limits_{h \to 0} \frac{f(x_1,\ldots,x_j+h,\ldots,x_n)-f(x_1,\ldots,x_n)}{h}$ existiert. Die Schreibweise dafür ist auch $\frac{\partial f(x)}{\partial x_j} = D_j f(x) = f_{x_j}(x)$. (Für $n = 2$ oder 3 werden die Variablen oft mit x, y, z bezeichnet.) Ist $\frac{\partial f}{\partial x_j} : A \to \mathbb{R}$ wieder partiell nach x_k differenzierbar, etc., bezeichnen wir diese **höheren Ableitungen** mit $\frac{\partial^2 f}{\partial x_k \partial x_j}(x) = D_k D_j f(x)$ etc.

Somit: f ist in x partiell nach x_j differenzierbar $\Leftrightarrow (x_j \mapsto f(x_1, \ldots, x_{j-1}, x_j, x_{i+1} \ldots x_n)$ ist differenzierbar.) Man berechnet also eine partielle Ableitung $\frac{\partial f}{\partial x_j}$, indem man alle anderen Variablen $x_0, \ldots, x_{j-1}, x_{j+1} \ldots x_n$ als Konstanten behandelt (also festhält) und nach x_j differenziert.



Beispiele.

1) $f : \mathbb{R}^2 \to \mathbb{R}$, $f(x,y) = xe^y + \sin(xy)$.
   $\frac{\partial f}{\partial x}(x,y) = e^y + \cos(xy) \cdot y$ (y fest), $\frac{\partial f}{\partial y}(x,y) = xe^y + \cos(xy)\ x$ (x fest).

2) $f : \mathbb{R}^3 \to \mathbb{R}$, $f(x,y,z) = x^2 + xy^2 + 2z^3$.
   $\frac{\partial f}{\partial x}(x,y,z) = 2x + y^2$, $\frac{\partial f}{\partial y}(x,y,z) = 2xy$, $\frac{\partial f}{\partial z}(x,y,z) = 6z^2$.

3) $f : \mathbb{R}^n - \{0\} \to \mathbb{R}$, $f(x) = \frac{1}{r}$, $r = \sqrt{x_1^2 + \cdots + x_n^2} = ||x||_2$, $x \neq 0$.
   $\frac{\partial f}{\partial x_j} = -x_j / r^3$, $j = 1 \ldots n$.
   Übung: $\frac{\partial r^\alpha}{\partial x_j}$.

4) f wie in 2). $\frac{\partial^2 f}{\partial y \partial x} = 2y$, $\frac{\partial^2 f}{\partial x \partial y} = 2y$.
   $\frac{\partial^2 f}{\partial x^2} = 2$, $\frac{\partial^2 f}{\partial y^2} = 2x$, $\frac{\partial^2 f}{\partial z^2} = 12z$, $\frac{\partial^2 f}{\partial z \partial y} = \frac{\partial^2 f}{\partial y \partial z} = 0$.

5) f wie in 3).
   $\frac{\partial^2 f}{\partial x_j^2} = \frac{3x_j^2}{r^5} - \frac{1}{r^3}$: $n = 3 \Rightarrow \sum\limits_{j=1}^{3} \frac{\partial^2 f}{\partial x_j^2} = \frac{3}{r^3} - \frac{3}{r^3} = 0$ $(x \neq 0)$.

7.2 Gemischte Ableitungen

In Beispiel 4) ergab sich $\frac{\partial^2 f}{\partial x \partial y} = \frac{\partial^2 f}{\partial y \partial x}$. Gilt dies allgemein?

Beispiel:

$f : \mathbb{R}^2 \to \mathbb{R}$, $f(x,y) = \begin{Bmatrix} xy \, \frac{x^2-y^2}{x^2+y^2} & (x,y) \neq 0 \\ 0 & (x,y) = 0 \end{Bmatrix}$.

Dann gilt

$$\frac{\partial f}{\partial y}(x,0) = \begin{Bmatrix} x & x \neq 0 \\ \lim\limits_{k\to 0} \frac{f(0,k)}{k} = 0 & x = 0 \end{Bmatrix} ,$$

$$\frac{\partial f}{\partial x}(0,y) = \begin{Bmatrix} -y & y \neq 0 \\ \lim\limits_{h\to 0} \frac{f(h,0)}{h} = 0 & y = 0 \end{Bmatrix} .$$

Diese beiden Funktionen haben stetige partielle Ableitungen, z.B. hat man $|\frac{\partial f}{\partial x}(0,y)| \leq 6\sqrt{y^2+y^2}$. Es ergibt sich $\frac{\partial f}{\partial x \partial y}(0,0) = 1 \neq -1 = \frac{\partial f}{\partial y \partial x}(0,0)$: Beide gemischten partiellen zweiten Ableitungen sind **unstetig** in $(0,0)$.

Definition. Sei $\Omega \subseteq \mathbb{R}^n$ offen, $\ell \in \mathbb{N}$. Wir definieren die C^ℓ**-Funktionen** auf Ω durch $C^\ell(\Omega) := \{f : \Omega \to \mathbb{R} \mid D_{j_\ell} \cdots D_{j_1} f : \Omega \to \mathbb{R}$ existiert für alle $j_i \in \{1, \cdots, n\}$ **und ist stetig**$\}$.

Satz (Schwarz)**.** *Sei $f \in C^\ell(\Omega)$. Dann sind alle partiellen Ableitungen der Ordnung $\leq \ell$ unabhängig von der Differentiationsreihenfolge.*

Beweis. Durch Induktion kann man o.B.d.A. annehmen, dass $\ell = 2$ ist. Dann sind nur 2 Variablen $\{x_j, x_k\}$ veränderlich, die anderen fest. Also nehmen wir zur Vereinfachung der Schreibweise o.B.d.A. $n = 2$ an. Sei also $f : \Omega \subseteq \mathbb{R}^2 \to \mathbb{R}$ in $C^2(\Omega)$. Zu zeigen: $\frac{\partial^2 f}{\partial y \partial x} = \frac{\partial^2 f}{\partial x \partial y}$. Sei $(x_0, y_0) \in \Omega$ und $\varepsilon > 0$ so klein, dass für alle $h, k \in \mathbb{R}$ mit $|h|,\ |k| < \varepsilon$ gilt: $(x_0+h, y_0+k) \in \Omega$. Die Funktion $\varphi : [x_0, x_0+h] \to \mathbb{R},\ \varphi(x) := f(x, y_0+k) - f(x, y_0)$ ist differenzierbar nach x . Nach dem Mittelwertsatz existiert $x_1 \in (x_0, x_0+h)$ mit $\varphi(x_0+h) - \varphi(x_0) = h\ \varphi'(x_1)$. Definiere $F : (0,\varepsilon)^2 \to \mathbb{R}$,

$$F(h,k) := f(x_0+h, y_0+k) - f(x_0+h, y_0) - f(x_0, y_0+k) + f(x_0, y_0) .$$

Dann ist $F(h,k) = \varphi(x_0+h) - \varphi(x_0) = h\ \varphi'(x_1) = h \left[\frac{\partial f}{\partial x}(x_1, y_0+k) - \frac{\partial f}{\partial x}(x_1, y_0)\right]$. Eine zweite Anwendung des Mittelwertsatzes ergibt einen Punkt $y_1 \in (y_0, y_0+k)$ mit

$$F(h,k) = hk \frac{\partial^2 f}{\partial y \partial x}(x_1, y_1) .$$

Analog definieren wir $\psi : [y_0, y_0+k] \to \mathbb{R},\ \psi(y) := f(x_0+h, y) - f(x_0, y)$. Ein ähnliches Argument zeigt:

$$F(h,k) = \psi(y_0+k) - \psi(y_0) = hk\ \frac{\partial^2 f}{\partial x \partial y}(x_2, y_2)$$

mit gewissen Punkten $x_2 \in (x_0, x_0+h),\ y_2 \in (y_0, y_0+k)$. Da $hk \neq 0$ ist, folgt

$$\frac{\partial^2 f}{\partial y \partial x}(x_1, y_1) = \frac{\partial^2 f}{\partial x \partial y}(x_2, y_2) \ .$$

Die Stetigkeit beider gemischten partiellen Abbildungen liefert für $|h|, |k| \to 0$, d.h. für $(x_1, y_1) \to (x_0, y_0) \leftarrow (x_2, y_2)$, dass $\frac{\partial^2 f}{\partial y \partial x}(x_0, y_0) = \frac{\partial^2 f}{\partial x \partial y}(x_0, y_0)$ gilt. $\square$

7.3 Änderung der C^1-Funktionen

Eine Funktion, deren partielle Ableitungen existieren, ist i.a. nicht einmal stetig: Etwa:
$f(x,y) = \left\{ \begin{array}{cc} x & y=0 \\ y & x=0 \\ \text{beliebig} & \text{sonst} \end{array} \right\}$ erfüllt $\frac{\partial f}{\partial x}(0) = \frac{\partial f}{\partial y}(0) = 1$, aber f ist **unstetig**!
Die partiellen Ableitungen von $f : \mathbb{R}^2 \to \mathbb{R}$ in (x_0, y_0) machen Aussagen über die Änderung von $f(x_0 + h, y_0) - f(x_0, y_0)$ bzw. $f(x_0, y_0 + k) - f(x_0, y_0)$. I.a. enthalten sie aber keine Information über "Gesamtänderung"

$$f(x_0 + h, y_0 + k) - f(x_0, y_0) \ .$$

Dafür benötigt man wieder die **Stetigkeit** von $\frac{\partial f}{\partial x}, \ \frac{\partial f}{\partial y}$ in x_0 (d.h. $f \in C^1$):

Satz 1. *Sei $\Omega \subseteq \mathbb{R}^n$ offen und $f \in C^1(\Omega), \ x \in \Omega$. Dann ist f stetig und für hinreichend kleines $\varepsilon > 0$ und $h \in \mathbb{R}^n$ mit $||h||_2 < \varepsilon$ gilt:*

$$f(x+h) - f(x) = \sum_{i=1}^{n} \frac{\partial f(x)}{\partial x_j} h_j + r(h) \quad \textit{mit} \quad \lim_{h \to 0} \frac{r(h)}{||h||_2} = 0 \ .$$

Beweis. Wähle $\varepsilon > 0$ so klein, dass aus $||y - x||_2 < \varepsilon$ folgt: $y \in \Omega$. Für $h = (h_1, \ldots, h_n)$ mit $||h||_2 < \varepsilon$ setzen wir (mit $e_i = i$-ter Einheitsvektor)

$$h^{(j)} = \sum_{i=1}^{j} h_i e_i \quad (j = 1, \ldots, n) \ ; \ h^{(0)} = 0, \ h^{(n)} = h \ .$$

Wegen $||h||_2 < \varepsilon$ ist $||h^{(j)}||_2 < \varepsilon$. Nach dem Mittelwertsatz gibt es Punkte $x^{(j)}$ auf der Strecke von $x + h^{(j-1)}$ nach $x + h^{(j)}$ (bei alleiniger Änderung in der Koordinate x_j), so dass

$$\begin{aligned} f(x+h) - f(x) &= \sum_{j=1}^{n} [f(x + h^{(j)}) - f(x + h^{(j-1)})] \\ &= \sum_{j=1}^{n} h_j \cdot \frac{\partial f}{\partial x_j}(x^{(j)}) = \sum_{j=1}^{n} \frac{\partial f}{\partial x_j}(x) \ h_j \ + r'(h) \end{aligned}$$

mit $r(h) := \sum_{j=1}^{n} \left(\frac{\partial f}{\partial x_j}(x^{(j)}) - \frac{\partial f}{\partial x_j}(x) \right) \cdot h_j$ gilt. Eine Anwendung der Ungleichung von Cauchy-Schwarz ergibt:

$$\frac{|r(h)|}{||h||_2} \leq \left(\sum_{j=1}^{n} \left| \frac{\partial f}{\partial x_j}(x^{(j)}) - \frac{\partial f}{\partial x_j}(x) \right|^2 \right)^{1/2} \ .$$

Nach Voraussetzung ist $\frac{\partial f}{\partial x_j}$ stetig in x . Für $h \to 0$ geht $x^{(j)} \to x$, somit geht die Summe gegen Null, d.h. $\lim\limits_{h\to 0} \frac{r(h)}{||h||_2} = 0$. Speziell ist f stetig. $\square$

Bemerkung. Man nennt $\nabla f : \Omega \to \mathbb{R}^n,\ \nabla f(x) = \left(\frac{\partial f}{\partial x_1}(x), \ldots, \frac{\partial f}{\partial x_n}(x)\right)$ auch den **Gradienten von f in x** . Damit hat man

$$f(x+h) - f(x) = \nabla f(x) \cdot h + r(h) \ .$$

Allgemeiner betrachten wir jetzt Abbildungen $f : \Omega \subseteq \mathbb{R}^n \to \mathbb{R}^m$ (z.B. ein räumliches Feld). Dann ist $f = \begin{pmatrix} f_1 \\ \vdots \\ f_m \end{pmatrix}$ mit $f_j : \Omega \to \mathbb{R}$.

Definition. Seien $\ell, n, m \in \mathbb{N}$ und $\Omega \subseteq \mathbb{R}^n$ offen. Sei $f : \Omega \to \mathbb{R}^m$. Dann ist $f \in C^\ell(\Omega, \mathbb{R}^m) :\Leftrightarrow$ Alle partiellen Ableitungen bis zur ℓ-ten Ordnung existieren und sind stetig. In Koordinatenschreibweise ist

$$\frac{\partial f}{\partial x_j} := {}^t\!\left(\frac{\partial f_1}{\partial x_j}, \ldots, \frac{\partial f_m}{\partial x_j}\right)$$

etc. Man nennt

$$f'(x_0) = Df(x_0) = \frac{df(x_0)}{dx} = \begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \vdots & & \vdots \\ \frac{\partial f_m}{\partial x_1} & \cdots & \frac{\partial f_m}{\partial x_n} \end{pmatrix} (x_0)$$

die **Funktionalmatrix** oder **Jacobimatrix von f in x_0** .

Satz 2. *Sei $\Omega \subseteq \mathbb{R}^n$ offen und $f : \Omega \to \mathbb{R}^m$ in $C^1(\Omega, \mathbb{R}^m)$. Sei $x_0 \in \Omega$. Dann gilt für alle $h \in \mathbb{R}^n - \{0\}$ mit kleiner Norm*

$$f(x+h) - f(x) = Df(x)\ h + r(h)$$

mit $r(h) \in \mathbb{R}^m$ und $\lim\limits_{h\to 0} \frac{r(h)}{||h||} = 0$, wobei der Limes im $\mathbb{R}^m$ gebildet wird.

Beweis. Für $j = 1, \ldots, m$ gilt nach Satz 1

$$f_j(x+h) - f_j(x) = \nabla f_j(x) \cdot h + r_j(h) \ , \ \lim_{h\to 0} \frac{r_j(h)}{||h||} = 0 \text{ (in } \mathbb{R}) \ .$$

Fasst man dies für alle $j = 1, \ldots, m$ zusammen, ergibt sich die Behauptung.
In Matrixschreibweise ist: $Df(x){\cdot}h = {}^t(\nabla f_1(x){\cdot}h, \ldots, \nabla f_m(x){\cdot}h)$, $r(h) = {}^t(r_1(h), \ldots, r_m(h))$.
$\square$

7.4 Differenzierbarkeit

Seien $f:\mathbb{R}\to\mathbb{R}$ bzw. $g:\mathbb{R}^n\to\mathbb{R}^m$ $\quad C^1$-Funktionen. Dann gilt

$$f(y_0+k)-f(k)=f'(y_0)k+R(k)\ ,\ g(x_0+h)-g(x_0)=g'(x_0)h+r(h)$$

mit $f'(y_0)\in\mathbb{R}$, $g'(x_0)\in\mathbb{R}^{m\times n}$ (Matrix oder lineare Abbildung: $\mathbb{R}^n\to\mathbb{R}^m$), $k\in\mathbb{R}$, $h\in\mathbb{R}^n$, $\lim\limits_{k\to 0}\frac{R(k)}{k}=0$, $\lim\limits_{h\to 0}\frac{r(h)}{||h||}=0$. Dies führt zu der folgenden Definition:

Definition. Seien X,Y Banachräume und $\Omega\subseteq X$ offen, ferner $x_0\in\Omega$. Eine Abbildung $f:\Omega\to Y$ heißt **differenzierbar in** x_0 $:\Leftrightarrow$. Es gibt eine stetige lineare Abbildung $T:X\to Y$ mit: $\exists\varepsilon>0\ \forall h\in X$, $||h||_X<\varepsilon$

$$f(x_0+h)-f(x_0)=Th+r(h)\ \text{ mit }\ \lim_{h\to 0}\frac{r(h)}{||h||}=0\ (\text{in } Y)\ . \tag{$*$}$$

Satz. *Sei $f:\Omega\subseteq\mathbb{R}^n\to\mathbb{R}^m$ differenzierbar in x_0 . Dann ist T eindeutig bestimmt, und jede Koordinatenfunktion f_i von f, $f=\begin{pmatrix}f_1\\ \vdots\\ f_m\end{pmatrix}$, ist partiell differenzierbar nach allen x_j $(j=1,\ldots,m)$ und T ist die Jacobi-Matrix $T=\begin{pmatrix}\frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_n}\\ \\ \frac{\partial f_m}{\partial x_1} & \cdots & \frac{\partial f_m}{\partial x_n}\end{pmatrix}(x_0)$, die automatisch stetig ist.*

Beweis. Mit $T=(t_{ij})_{i=1,\cdots,m,\ j=1,\cdots,n}$ ist

$$f_i(x_0+h)-f_i(x_0)=\sum_{j=1}^n t_{ij}h_j+r_i(h)\ ,\ \lim_{h\to 0}\frac{r_i(h)}{||h||}=0\ .$$

Wähle $h=h_ke_k\in\mathbb{R}^n$. Dann besitzt

$$\begin{aligned} f_i(x_0+h)-f_i(x_0) &= t_{ik}h_k+r_i(h)\ ,\\ \frac{f_i(x_0+h)-f_i(x_0)}{h_k} &= t_{ik}+\frac{r_i(h)}{h_k}\end{aligned}$$

einen Limes für $h\to 0$, d.h. für $h_k\to 0$, und es ist $t_{ik}=\frac{\partial f_i}{\partial x_k}(x_0)$. Also ist f_i partiell nach x_k differenzierbar und t_{ik} ist damit eindeutig bestimmt. $\square$

Vergleicht man den Satz mit Satz 2 aus 7.3, rechtfertigt dies die folgende Bezeichnung.

Definition. $f:\Omega\subseteq X\to Y$ sei differenzierbar in $x_0\in\Omega$. Dann heißt die stetige lineare Abbildung T aus $(*)$ die **(totale) Ableitung von** f **in** x_0 und wird mit $Df(x_0)=f'(x_0)=\frac{df}{dx}(x_0)$ bezeichnet. Im Fall $X=\mathbb{R}^n,\ Y=\mathbb{R}^m$ ergibt sich gerade die Funktional- oder Jacobimatrix (in Matrixdarstellung).

Bemerkungen:

1) Der Begriff ist im allgemeinen von Normen auf X und Y abhängig. Im Falle $X = \mathbb{R}^n,\ Y = \mathbb{R}^m$ jedoch nicht, da alle Normen auf X bzw. auf Y äquivalent sind.

2) Die Ableitung von $f : \Omega \subseteq X \to Y$ ist also eine Abbildung

$$Df : \Omega \to L(X, Y) \ .$$

3) $Y = \mathbb{R}\ :\ Df(x_0) : \mathbb{R}^n \to \mathbb{R}$ wird durch das Skalarprodukt mit dem **Gradienten** von f in x_0 gegeben, d.h. $Df(x_0)(v) =< \nabla f(x_0), v >$. Der Gradient spannt den **Tangentialraum** in $f(x_0)$ auf.

7.5 Stetigkeit und Differenzierbarkeit

Definition. Seien X, Y Banachräume, $\Omega \subseteq X$ offen und $x_0 \in \Omega$. Eine Abbildung $f : \Omega \to Y$ heißt **stetig differenzierbar in** Ω: $\Leftrightarrow$
$Df : \Omega \to L(X, Y)$ existiert und ist stetig in Ω . Wir schreiben $f \in C^1(\Omega, Y)$.
Dies stimmt mit Definition 7.2 überein, da gilt:

Satz. (a) *f ist differenzierbar (in Ω) $\Rightarrow$ f ist stetig (in Ω).*

(b) *Für $X = \mathbb{R}^n,\ Y = \mathbb{R}^m$ gilt: f ist stetig differenzierbar in $\Omega \Leftrightarrow f \in C^1(\Omega, \mathbb{R}^m)$.*

Beweis.

(a) Für $h \to 0$ geht $f(x_0 + h) - f(x_0) = Th + r(h)$ gegen Null. Da Konvergenz in $\mathbb{R}^{nm}$ koordinatenweise Konvergenz bedeutet, gilt mit 7.3, Satz 2, 7.4 und 7.2:
$Df : \Omega \to L(\mathbb{R}^n, \mathbb{R}^m) \simeq \mathbb{R}^{n \times m}$ ist stetig in x_0 für alle $x_0 \in \Omega \Leftrightarrow$

$$\forall j = 1, \dots, n, i = 1, \dots, m \qquad \frac{\partial f_i}{\partial x_j} = (Df)_{ij} \text{ ist stetig in } x_0 \text{ für alle } x_0 \in \Omega$$
$$\Leftrightarrow f \in C^1(\Omega, \mathbb{R}^m) \ .$$

Wir haben folgende Implikationen:

f ist stetig differenzierbar $\overset{\not\Leftarrow}{\Rightarrow} f$ ist differenzierbar
$\overset{\not\Leftarrow}{\Rightarrow}$ Alle 1. partiellen Ableitungen von f existieren
$\overset{\not\Leftarrow}{\Rightarrow}$ f ist stetig

Übungen:

1) Eigenschaften durch Beispiele abgrenzen.

2) $f : \Omega \subseteq \mathbb{R}^n \to \mathbb{R}^m$ ist differenzierbar in $x_0 \iff f_1, \dots, f_m$ differenzierbar in x_0 .

Beispiele:

a) Für konstante Funktionen $f : \mathbb{R}^n \to \mathbb{R}^m\ ,\ x \mapsto c \in \mathbb{R}^m$ gilt: $Df(x) = 0,\ x \in \mathbb{R}^n$.

b) Sei $f : X \to Y$ stetig und linear, d.h. $f(x) = T(x)$ mit $T \in L(X, Y)$. Da

$$f(x_0 + h) - f(x_0) = T(x_0 + h) - T(x_0) = T(h) = T(h) + 0\ ,\ r(h) = 0\ ,$$

folgt $f'(x_0) = T = f$ für alle $x_0 \in X$. Somit ist $f'(x_0) = F : X \to Y$, lineare Abbildungen reproduzieren sich selbst bei Ableitung in einem festen Punkt.

c) Sei $f : (\mathbb{R}_+)^2 \to \mathbb{R}^2,\ f(x, y) = \begin{pmatrix} x + \sqrt{y} \\ \sqrt{x} + y \end{pmatrix}$. f ist differenzierbar mit

$$f'(x, y) = Df(x, y) = \begin{pmatrix} 1 & \frac{1}{2\sqrt{y}} \\ \frac{1}{2\sqrt{x}} & 1 \end{pmatrix}\ .$$

d) $f : \mathbb{R}^n \to \mathbb{R}$ heißt **homogen vom Grad** $\alpha :\Leftrightarrow f(tx) = t^\alpha f(x)\ ;\ x \in \mathbb{R}^n,\ t \in \mathbb{R}^+$. Ist f differenzierbar, gilt $Df(x)x = \alpha f(x)$.

e) Die Polarkoordinaten-Abbildung $f : (0, \infty) \times (0, 2\pi) \to \mathbb{R}^2,\ (x, y) \mapsto (r\cos\varphi, r\sin\varphi)$ hat die Jacobi-Matrix $J = \begin{pmatrix} \cos\varphi & -r\sin\varphi \\ \sin\varphi & r\cos\varphi \end{pmatrix}$ mit $\det(J) = r \neq 0$ für $(x, y) \neq 0$.

7.6 Differentiationsregeln

Satz 1. *Seien X, Y Banachräume und $\Omega \subseteq X$ offen. Seien $f, g : \Omega \to Y$ differenzierbar in $x_0 \in \Omega$. Dann gilt:*

(a) $f + g, \alpha f$ sind differenzierbar in x_0 mit $(f + g)'(x_0) = f'(x_0) + g'(x_0)$, $(\alpha f)'(x_0) = \alpha f'(x_0)$.

(b) Für $Y = \mathbb{R}$ ist $f \cdot g$ differenzierbar in x_0 mit $(fg)'(x_0) = f(x_0)g'(x_0) + g(x_0)f'(x_0)$. (Als Elemente von $X^* = L(X, \mathbb{R})$; $f(x_0), g(x_0) \in \mathbb{R}$.) Analog für den Quotienten $f/g, g(x_0) \neq 0$.

Beweis.

(a) $\left.\begin{matrix} f(x_0+h) - f(x_0) = f'(x_0)h + r_1(h) \\ g(x_0+h) - g(x_0) = g'(x_0)g + r_2(h) \end{matrix}\right\}\qquad \frac{r_i(h)}{||h||} \to 0\ (h \to 0, i = 1, 2)$

$\Rightarrow (f + g)(x_0 + h) - (f + g)(x_0) = (f'(x_0) + g'(x_0))(h) + r(h),\ r(h) := r_1(h) + r_2(h)$.

Also existiert $(f + g)'(x_0)$ und ist gleich $f'(x_0) + g'(x_0)$.

(b) Übung.

$\square$

Satz 2. (Kettenregel). *Seien X, Y, Z Banachräume, $F \subseteq X$ offen, $G \subseteq Y$ offen und $f : F \to Y$ mit $f(F) \subseteq G$, $g : G \to Z$. Ist dann f in $x_0 \in F$ differenzierbar, g in $y_0 := f(x_0) \in G$ differenzierbar, ist $g \circ f : F \to Z$ in x_0 differenzierbar mit*

$$(g \circ f)'(x_0) = g'(y_0) \circ f'(x_0) \ .$$

Bemerkung: Letzteres ist als Operatorenhintereinanderschaltung zu interpretieren mit

$$f'(x_0) \in L(X, Y),\ g'(y_0) \in L(Y, Z) \ .$$

Beweis: Für kleine $\varepsilon > 0$ und alle $h \in X$ mit $||h||_X < \varepsilon$ existiert nach Voraussetzung $(g \circ f)(x_0 + h)$. Da g differenzierbar ist, gilt

$$\begin{aligned}(g \circ f)(x_0 + h) - (g \circ f)(x_0) &= g(f(x_0 + h)) - g(f(x_0)) \\ &= g'(f(x_0))[f(x_0 + h) - f(x_0)] + r_1(f(x_0 + h) - f(x_0))\end{aligned}$$

mit $\lim\limits_{k \to 0} \frac{r_1(k)}{||k||} = 0$ in Y . Setze $\varrho(k) := \frac{r_1(k)}{||k||}$, $k \neq 0$. Also: $\lim\limits_{k \to 0} \varrho(k) = 0$ in Y . Da f differenzierbar ist, gilt $f(x_0 + h) - f(x_0) = f'(x_0) \cdot h + r_2(h)$ mit $\lim\limits_{h \to 0} \frac{r_2(h)}{||h||} = 0$ in X . Es folgt:

$$\begin{aligned}(g \circ f)(x_0 + h) - (g \circ f)(x_0) &= g'(f(x_0))[f'(x_0)h + r_2(h)] + ||f'(x_0)h + r_2(h)||\ \varrho(f'(x_0)h + r_2(h)) \\ &= g'(f(x_0)) \circ f'(x_0)\ h + r(h) \ ,\end{aligned}$$

wobei

$$r(h) := g'(f(x_0))\ r_2(h) + ||f'(x_0)h + r_2(h)||\ \varrho(f'(x_0)h + r_2(h))$$

d.h. $\frac{r(h)}{||h||} \underset{h \to 0}{\longrightarrow} 0$, da wegen der Stetigkeit der linearen Abbildungen $g'(f(x_0))$ und $f'(x_0)$ gilt:

$$g'(f(x_0) \left(\underbrace{\frac{r_2(h)}{||h||}}_{\to 0} \right) \to 0,\ \varrho(\underbrace{f'(x_0)h}_{\to 0} + \underbrace{r_2(h)}_{\to 0}) \to 0 \ ,$$

und $||f'(x_0)\frac{h}{||h||} + \frac{r_2(h)}{||h||}|| \leq ||f'(x_0)|| + 1$ beschränkt ist, da für kleine h gilt $||r_2(h)|| \leq ||h||$. Es folgt also

$$(g \circ f)'(x_0) = g'(y_0) \circ f'(x_0) \ .$$

□

Spezialfälle im $\mathbb{R}^n$:

Seien $X = \mathbb{R}^n, Y = \mathbb{R}^m, Z = \mathbb{R}^\ell, f = \begin{pmatrix} f_1 \\ \vdots \\ f_m \end{pmatrix}$, $g = \begin{pmatrix} g_1 \\ \vdots \\ g_\ell \end{pmatrix}$ wie in Satz 2 mit $F \subseteq X$, $G \subseteq Y$. Bezeichnen x_j die Variablen in X und y_k die Variablen in Y, lässt sich die Kettenregel in Matrixschreibweise so formulieren:

$$
\begin{aligned}
(g\circ f)(x) &= \begin{pmatrix} g_1(f_1(x),\ldots,f_m(x)) \\ \vdots \\ g_\ell(f_1(x),\ldots,f_m(x)) \end{pmatrix} = \begin{pmatrix} (g\circ f)_1(x) \\ \vdots \\ (g\circ f)_\ell(x) \end{pmatrix} , \\
D(g\circ f)(x_0) &= \begin{pmatrix} \frac{\partial (g\circ f)_1}{\partial x_1} & \cdots & \frac{\partial (g\circ f)_1}{\partial x_n} \\ \vdots & & \vdots \\ \frac{\partial (g\circ f)_\ell}{\partial x_1} & \cdots & \frac{\partial (g\circ f)_\ell}{\partial x_n} \end{pmatrix}(x_0) = \begin{pmatrix} \frac{\partial g_1}{\partial y_1} & \cdots & \frac{\partial g_1}{\partial y_m} \\ \vdots & & \vdots \\ \frac{\partial g_\ell}{\partial y_1} & \cdots & \frac{\partial g_\ell}{\partial y_m} \end{pmatrix}(f(x_0)) \circ \begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \vdots & & \vdots \\ \frac{\partial f_m}{\partial x_1} & \cdots & \frac{\partial f_n}{\partial x_n} \end{pmatrix}(x_0) .
\end{aligned}
$$

Das bedeutet

$$
\frac{\partial (g\circ f)_i}{\partial x_j}(x_0) = \sum_{k=1}^{m} \frac{\partial g_i}{\partial y_k}(y_0)\ \frac{\partial f_k}{\partial x_j}(x_0) \ \text{ für } \ i=1,\ldots,\ell,\ j=1,\ldots,n_0 . \tag{1}
$$

Speziell ist der Fall $\ell = 1$ wichtig, $g\circ f \ : \ \mathbb{R}^n \to \mathbb{R}$. Dann ist

$$
\frac{\partial (g\circ f)}{\partial x_j}(x_0) = \sum_{k=1}^{m} \frac{\partial g}{\partial y_k}(y_0)\frac{\partial f_k}{\partial x_j}(x_0) .
$$

Für $n = m = \ell$ erhält man quadratische Matrizen; die Determinante der Jacobi-Matrix wird sich später als wichtig herausstellen; man hat dann

$$
\det\left(D(g\circ f)(x_0)\right) = \det\left(Dg(y_0)\right)\cdot\det\left(Df(x_0)\right) . \tag{2}
$$

Beispiel: Sei $f : (\mathbb{R}_+)^2 \to (\mathbb{R}_+)^2$, $f_1(x_1,x_2) = x_1x_2$, $f_2(x_1,x_2) = \sqrt{x_1}/x_2$
$g : \mathbb{R}^2 - \{0\} \to \mathbb{R}$, $g(y_1,y_2) = \ln(y_1^2+y_2^2)$.
Dann ist $(g\circ f)(x_1,x_2) = \ln(x_1^2x_2^2 + x_1/x_2^2) =: h(x_1,x_2)$,

$$
\begin{aligned}
\frac{\partial h}{\partial x_1} &= \frac{\partial g}{\partial y_1}\ \frac{\partial f_1}{\partial x_1} + \frac{\partial g}{\partial y_2}\ \frac{\partial f_2}{\partial x_1} = \frac{2y_1}{y_1^2+y_2^2}\cdot x_2 + \frac{2y_2}{y_1^2+y_2^2}\cdot \frac{1}{2\sqrt{x_1}x_2}\Bigg|_{y_i=f_i(x_1,x_2)} \\
&= \frac{2x_1x_2^4+1}{x_1^2x_2^4+x_1} \ \text{ (dies ist natürlich auch direkt berechenbar!)} , \\
\frac{\partial h}{\partial x_2} &= \frac{\partial g}{\partial y_1}\ \frac{\partial f_1}{\partial x_2} + \frac{\partial g}{\partial y_2}\ \frac{\partial f_2}{\partial x_2} = \frac{2y_1}{y_1^2+y_2^2}\cdot x_1 + \frac{2y_2}{y_1^2+y_2^2}\cdot \left(-\frac{\sqrt{x_1}}{x_2^2}\right)\Bigg|_{y_i=f_i(x_1,x_2)} \\
&= \frac{2(x_1^2x_2^3 - x_1/x_2)}{x_1^2x_2^4+x_1} .
\end{aligned}
$$

7.7 Richtungsableitungen

Man interessiert sich nicht nur für Funktionsänderungen in koordinatenparallelen Richtungen x_j , sondern in beliebigen Richtungen. Das führt zum Begriff der Richtungsableitung, der die partiellen Ableitungen verallgemeinert.

Definition. Sei $x_0 \in \Omega \subseteq \mathbb{R}^n$ offen, $f : \Omega \to \mathbb{R}^m$. Sei $v \in \mathbb{R}^n, ||v||_2 = 1$. Man nennt $\frac{\partial f}{\partial v}(x_0) = \lim\limits_{t\to 0} \frac{f(x_0+tv)-f(x_0)}{t}$ die **(Richtungs-)Ableitung von f in x_0 in Richtung** v, falls der Limes existiert.

Offenbar ist $\frac{\partial f}{\partial x_j} = \frac{\partial f}{\partial e_j}$. O.B.d.A. sei $m = 1$, sonst wende man den folgenden Satz auf die Koordinatenfunktionen an.

Satz. *Sei $\Omega \subseteq \mathbb{R}^n$ offen und $x_0 \in \Omega$. Sei $f : \Omega \to \mathbb{R}$ in x_0 differenzierbar. Dann existiert $\frac{\partial f}{\partial v}(x_0)$ für alle Richtungsvektoren $v \in \mathbb{R}^n,\ ||v||_2 = 1$ und es gilt*

$$\frac{\partial f}{\partial v}(x_0) = f'(x_0)(v) = \nabla f(x_0) \cdot v = \sum_{i=1}^{n} \frac{\partial f(x_0)}{\partial x_i} v_i \ . \tag{1}$$

Falls $\nabla f(x_0) \neq 0$ ist, gibt es unter allen Richtungsableitungen $\frac{\partial f}{\partial v}(x_0)$ – v variierend mit $||v||_2 = 1$ – eine größte, nämlich die Ableitung in Gradientenrichtung $v = \frac{\nabla f(x_0)}{||\nabla f(x_0)||_2}$ mit dem Wert $||\nabla f(x_0)||_2$.

Beweis.

(i) Es ist $\frac{f(x_0+tv)-f(x_0)}{t} = \frac{f'(x_0)(tv)+r(tv)}{t} = f'(x_0)(v) + \frac{r(tv)}{t}$. Da $\lim\limits_{t\to 0} \frac{r(tv)}{t} = 0$, folgt die erste Behauptung (1). Für $v = (v_i)_{i=1}^n \in \mathbb{R}^n$ mit $||v||_2 = 1$ gilt

(ii) $\left|\frac{\partial f}{\partial v}(x_0)\right| \leq \left(\sum\limits_{i=1}^{n} \left| \frac{\partial f(x_0)}{\partial x_i} \right|^2 \right)^{1/2} ||v||_2$ mit der Ungleichung von Cauchy-Schwarz, also

$\left|\frac{\partial f}{\partial v}(x_0)\right| \leq ||\nabla f(x_0)||_2$. Speziell für $v_0 = \frac{\nabla f(x_0)}{||\nabla f(x_0)||_2}$ folgt aus (1)

$$\frac{\partial f}{\partial v_0}(x_0) = \frac{\nabla f(x_0) \cdot \nabla f(x_0)}{||\nabla f(x_0)||_2} = ||\nabla f(x_0)||_2 \ ,$$

d.h. das Maximum von $\frac{\partial f}{\partial v}(x_0)$ für alle obigen v wird für die Gradientenrichtung v_0 angenommen.

□

Für die entgegengesetzte Richtung $(-v_0)$ ist $\frac{\partial f}{\partial(-v_0)}(x_0) = -\frac{\partial f}{\partial v_0}(x_0) = -||\nabla f(x_0)||_2$: die Richtung des Gradienten ist die Richtung stärksten Anstiegs von f , die Richtung des negativen Gradienten die Richtung des stärksten Abfallens von f (nahe x_0).

7.8 Der Mittelwertsatz

Satz. *(Mittelwertsatz) Sei X ein Banachraum und sei $\Omega \subseteq X$ offen. Sei $f : \Omega \to \mathbb{R}$ stetig differenzierbar. Dann gilt für je zwei Punkt $x_0,\ x \in \Omega$, für die die verbindende Strecke S ganz in Ω liegt: Es gibt $\theta, 0 < \theta < 1$ mit $f(x) = f(x_0) + Df(x_0 + \theta(x - x_0))(x - x_0)$.*

Beweis. Definiere $\varphi : [0,1] \to \mathbb{R},\ \varphi(t) := f(x_0 + t(x - x_0))$. φ ist nach Voraussetzung wohldefiniert und differenzierbar. Nach der Kettenregel gilt:

$$\frac{d\varphi}{dt}(t) = Df(x_0 + t(x - x_0)) \cdot (x - x_0) \ , \text{ da } \frac{d}{dt}(x_0 + t(x - x_0)) = x - x_0 \text{ ist.}$$

Der Mittelwertsatz für reelle Funktionen einer Veränderlichen mit $\varphi(0) = f(x_0),\ \varphi(1) = f(x)$, liefert dann die Behauptung. □

Bemerkung. Für $X = \mathbb{R}^n$ ist $Df(x) = \nabla f(x),\ f(x) = f(x_0) + \nabla f(x_0 + \theta(x - x_0)) \cdot (x - x_0)$, wobei $\cdot$ das Skalarprodukt im $\mathbb{R}^n$ bezeichnet. Versteht man die folgenden Integrationen komponentenweise, so ist es richtig, dass

$$\begin{aligned} f(x) - f(x_0) &= \varphi(1) - \varphi(0) = \int_0^1 \frac{d\varphi}{dt}(t)dt \\ &\overset{(*)}{=} \left(\int_0^1 \nabla f(x_0 + t(x - x_0))dt \right) \cdot (x - x_0). \end{aligned}$$

Eine Anwendung auf die Komponentenfunktionen von $f : \Omega \subseteq \mathbb{R}^m$ liefert das folgende

Korollar. Sei $f : \Omega \subseteq \mathbb{R}^n \to \mathbb{R}^m$ stetig differenzierbar und seien $x_0, x \in \Omega$ mit Verbindungsstrecke S ganz in Ω . Dann gilt

$$||f(x) - f(x_0)|| \leq M||x - x_0|| \ , \ M = \sup_{y \in S} ||f'(y)||_{op} < \infty \ .$$

Beweis. Aus $(*)$ und der Dreiecksungleichung, angewandt auf die Riemann-Summen im Integral, folgt im Grenzwert, dass

$$\begin{aligned} ||f(x) - f(x_0)|| &= || \left(\int_0^1 \nabla f_i(x_0 + t(x - x_0)) \cdot (x - x_0)dt \right)_{i=1}^m || \\ &\leq \int_0^1 ||(\nabla f_i(x_0 + t(x - x_0)) \cdot (x - x_0))_{i=1}^m||dt \\ &\leq \int_0^1 ||Df(x_0 + t(x - x_0))||_{op}||x - x_0||dt \\ &\leq \left(\sup_{y \in S} ||Df(y)||_{op} \right) \ ||x - x_0|| \ . \end{aligned}$$

Nun ist $||Df(\cdot)||_{op} \ : S \xrightarrow{Df} L(\mathbb{R}^n, \mathbb{R}^m) \xrightarrow{||\cdot||_{op}} \mathbb{R}$ stetig, nimmt also sein Maximum auf S an; daher ist $M = \sup_{y \in S} ||Df(y)||_{op} < \infty$. □

Bemerkung: In den meisten Fällen wird der Mittelwertsatz in einer Ungleichungsform wie im Korollar verwandt.

7.9 Die Taylorsche Formel

Wendet man die Taylorsche Formel für Funktionen in $\mathbb{R}$ auf die Funktion φ aus 7.8 an, ergibt sich die Taylorformel im $\mathbb{R}^n$. Dazu folgende Vorbemerkungen und
Notationen:

(1) Seien X, Y Banachräume, $\Omega \subseteq X$ offen und sei $f : \Omega \to Y$ zweimal differenzierbar. Dann bilden ab:

$$Df : \Omega \subseteq X \to L(X, Y) \text{ und } D^2 f : \Omega \subseteq X \to L(X, L(X, y))$$

Nun kann man aber $L(X, L(X, Y))$ mit den **bilinearen Abbildungen** von $X \times X$ nach Y identifizieren, $L(X, L(X, Y)) = L(X \times X, Y)$; allgemein für ℓ-mal differenzierbare Funktionen, die ℓ-fach multilinearen Abbildungen $L(X \times \cdots \times X, Y)$ betrachten: $D^\ell f : \Omega \subseteq X \longrightarrow L(\underbrace{X \times \cdots \times X}_{\ell}, Y)$.

Wir schreiben $D^\ell f(x_0)(x_1, \ldots, x_\ell)$ für die Anwendung von $D^\ell f(x_0)$ auf das n-Tupel $(x_1, \ldots, x_\ell)$.

(2) **Multiindizes**-Schreibweise: Sei $\alpha = (\alpha_1, \ldots, \alpha_n) \in \mathbb{N}^n$. Setze $|\alpha| = \sum\limits_{j=1}^{n} \alpha_j$ und für $\Omega \subseteq \mathbb{R}^n,\ f \in C^\ell(\Omega, Y)$ und $|\alpha| \leq \ell\ :\ D_\alpha f(x_0) := \frac{\partial^{|\alpha|} f(x_0)}{(\partial x_n)^{\alpha_n} \ldots (\partial x_1)^{\alpha_1}},\ D_\alpha f \in C(\Omega)$. $\alpha! := \alpha_1! \ldots \alpha_n!$, $x^\alpha := x_1^{\alpha_1} \ldots x_n^{\alpha_n}$ für $x \in \mathbb{R}^n$.

Satz (Taylorsche Formel)**.** *Sei X ein Banachraum und $\Omega \subseteq X$ offen. Sei $f : \Omega \to \mathbb{R}$ in $C^{k+1}(\Omega)$ und seien $x_0, x \in \Omega$ so, dass die Verbindungsstrecke S von x_0 nach x ganz in Ω liegt. Dann gibt es ein θ , $0 < \theta < 1$, mit*

$$f(x) = f(x_0) + \frac{Df(x_0)}{1!}(h) + \frac{D^2 f(x_0)}{2!}(h, h) + \cdots + \frac{D^k f(x_0)}{k!} \underbrace{(h, \ldots, h)}_{k}$$

$$+ \frac{D^{k+1} f(x_0 + \theta h)}{(k+1)!} \underbrace{(h, \ldots, h)}_{k+1} \text{ mit } h := x - x_0 \ .$$

Korollar. *Sei $\Omega \subseteq \mathbb{R}^n$ offen und $f \in C^{k+1}(\Omega)$. Seien $x_0, x \in \Omega$ mit Verbindungsstrecke S in Ω . Dann gibt es $\theta, 0 < \theta < 1$, mit*

$$f(x) = \sum_{|\alpha|=0}^{k} \frac{D_\alpha f(x_0)}{\alpha!} h^\alpha + \sum_{|\alpha|=k+1} \frac{D_\alpha f(x_0 + \theta h)}{\alpha!} h^\alpha \ , \ h := x - x_0 \ .$$

Beweis.

(1) Sei $h := x - x_0$. Setze $\varphi : [0, 1] \to \mathbb{R}$, $\varphi(t) := f(x_0 + th)$. Es folgt $\frac{d\varphi}{dt}(t) = Df(x_0 + th)(h)$. Die Kettenregel 7.6 liefert für die zweite Ableitung von φ

$$\frac{d^2 \varphi}{dt^2}(t) = D^2 f(x_0 + th)(h, h) \ .$$

Analog gilt

$$\frac{d^j \varphi}{dt^j}(t) = D^j f(x_0 + th) \underbrace{(h, \ldots, h)}_{j\text{-mal}} \ .$$

Somit liefert die Taylorsche Formel 5.8 in einer Variablen

$$\begin{aligned} f(x) &= \varphi(1) = \varphi(0) + \frac{\varphi'(0)}{1!} + \cdots + \frac{\varphi^{(k)}}{k!} + \frac{\varphi^{(k+1)}(\theta)}{(k+1)!} \\ &= f(x_0) + \frac{Df(x_0)}{1!}(h) + \cdots + \frac{D^k f(x_0)}{k!} \underbrace{(h, \ldots, h)}_{k} + \frac{D^{k+1} f(x_0 + \theta h)}{(k+1)!} \underbrace{(h, \ldots, h)}_{(k+1)} \ . \end{aligned}$$

(2) Im Fall des Korollars $X = \mathbb{R}^n$ benutzen wir die explizitere Darstellung der Ableitungen über partiellen Ableitungen (siehe Kettenregel):

$$\frac{d\varphi}{dt} = \nabla f(x_0 + th) \cdot h = (h \cdot \nabla) f(x_0 + th) \ .$$

Wendet man dieses auf $\frac{d\varphi}{dt}$ statt φ an, erhält man formal durch Induktion

$$\frac{d^j \varphi}{dt^j}(t) = (h \cdot \nabla)^{(j)} f(x_0 + th) \ , \ (f \cdot \nabla)^{(j)} f := (h \cdot \nabla)(h \cdot \nabla)^{j-1} f \ .$$

Wie beim binomischen (polinomischen) Lehrsatz gilt die Entwicklung

$$\begin{aligned} (h \cdot \nabla)^{(j)} &= (h_1 \frac{\partial}{\partial x_1} + \cdots + h_n \frac{\partial}{\partial x_n})^{(j)} \\ &= \sum_{|\alpha| = j} \frac{j!}{\alpha!} \ h^\alpha \ D_\alpha \ , \end{aligned}$$

somit ist

$$f(x) = \sum_{j=0}^{k} \sum_{|\alpha|=j} \frac{D_\alpha f(x_0)}{\alpha!} \ h^\alpha + \sum_{|\alpha|=k+1} \frac{D_\alpha f(x_0 + h\theta)}{\alpha!} h^\alpha = \sum_{|\alpha|=0}^{k} \cdots + \sum_{|\alpha|=k+1} \cdots$$

$\square$

Beispiel: Für $n = 2$ ist in abgekürzter Schreibweise (ohne Variablen)

$$\begin{aligned} &f(x_0 + h, y_0 + k) \\ &\quad = f + (f_x h + f_y h) + \frac{1}{2}(f_{xx} h^2 + 2 f_{xy} hk + f_{yy} h^2) + \frac{1}{6}(f_{xxx} h^3 + 3 f_{xxy} h^2 h + 3 f_{xyy} h h^2 + f_{yyy} h^2) + \cdots \end{aligned}$$

7.10 Taylorreihen und Potenzreihen

Das Problem der Konvergenz von Taylorreihen ist analog zum Eindimensionalen. Man hat Reihen über abzählbare Indexmengen J ohne natürliche Anordnung in einer Einfachreihe (bzgl. $\mathbb{N}$). Die Konvergenz kann daher sinnvollerweise als “unbedingte Konvergenz” untersucht werden:

Definition. Sei J abzählbar und seien $a_i \in \mathbb{R}$ für $i \in J$. Eine Reihe $\sum_{i \in J} a_i$ heißt **konvergent gegen** $a \in \mathbb{R}$ $:\Leftrightarrow$

$$\forall \varepsilon > 0 \ \exists \underset{\text{endlich}}{I_0 \subseteq J} \quad \forall I \subseteq J,\ I_0 \subseteq I \qquad \left| \sum_{i \in I} a_i - a \right| < \varepsilon \ .$$

Satz 1. $\sum\limits_{i \in J} a_i$ *konvergent* $\Leftrightarrow \sup\limits_{\substack{I \subseteq J \\ endlich}} \sum\limits_{i \in I} |a_i| < \infty$.

Die linke Aussage entspricht der unbedingten, die rechte der absoluten Konvergenz. In $\mathbb{N}$ sind beide nach 2.12 äquivalent, der Beweis dort lässt sich für diesen Fall modifizieren. Wir beweisen Satz 1 daher nicht (siehe Grauert-Lieb II, Kap. 3, §5). Wegen der absoluten Konvergenz sind die Voraussetzungen des großen Umordnungssatzes 2.13 erfüllt, man hat analog

Satz 2. *Sei* $\sum\limits_{i \in J} a_i = a$ *konvergent. Sei* $J = \bigcup\limits_{\lambda \in L} J_\lambda$, *L abzählbar,* J_λ *paarweise disjunkt. Mit* $\sum\limits_{i \in J_\lambda} a_i = a_\lambda$ *gilt:* $\sum\limits_{\lambda \in L} a_\lambda = a$ *konvergiert gegen* a .
(Dies entspricht $\sum\limits_n \sum\limits_m a_{nm} = \sum\limits_m \sum\limits_n a_{nm}$).

Satz 3. *Seien* $a_\alpha \in \mathbb{R}$ *für* $\alpha \in \mathbb{N}_0,\ x_0 \in \mathbb{R}^n$. *Es sei die Potenzreihe* $f(x) := \sum\limits_{|\alpha|=0}^{\infty} a_\alpha (x - x_0)^\alpha$ *in* $x_1 \in \mathbb{R}^n$ *mit* $c_j := |x_{1j} - x_{0j}| > 0$ $(\forall j = 1, \ldots, n)$ *konvergent im gerade definierten Sinne. Dann konvergiert die Potenzreihe auch für alle x im offenen Quader* $Q = \{x \in \mathbb{R}^n \mid |x_j - x_{0j}| < c_j, j = 1, \ldots, n\}$. *Sie ist dort beliebig oft differenzierbar, die Ableitungen berechnen sich durch gliedweise Differentiation und man hat* $D_\alpha f(x_0) = \alpha!\ a_\alpha$.

Beweis.

(a) O.B.d.A. sei $x_0 = 0$. Aus "$\sum\limits_{|\alpha|=0}^{\infty} a_\alpha x_1^\alpha$ konvergiert" folgt:

$$\exists R > 0 \ \forall \alpha \in \mathbb{N}_0^n \quad |a_\alpha x_1^\alpha| \leq R \ .$$

Sei $I_0 \subseteq \mathbb{N}_0^n$ endlich. Für $x \in Q,\ |x_j| < |x_{1j}|$ gilt dann

$$\sum_{\alpha \in I_0} |a_\alpha x^\alpha| = \sum_{\alpha \in I_0} |a_\alpha x_1^\alpha| \, \frac{|x^\alpha|}{|x_1^\alpha|} \leq R \sum_{\alpha \in I_0} q^\alpha$$

mit $q = (q_1, \ldots, q_n),\ q_i := \left| \frac{x_i}{x_{1i}} \right| < 1$. Aber die letztere geometrische Reihe konvergiert nach Satz 2,

$$\sum_{\alpha \in \mathbb{N}_0^n} q^\alpha = \sum_{\alpha_1=0}^{\infty} \cdots \sum_{\alpha_n=0}^{\infty} a_1^{\alpha_1} \ldots q_n^{\alpha_1} = \prod_{i=1}^{n} \frac{1}{1 - q_i} \ ,$$

also $\sup\limits_{I_0} \sum_{\alpha \in I_0} |a_\alpha x^\alpha| < \infty$: Die Reihe $\sum\limits_{\alpha \in \mathbb{N}_0^n} a_\alpha x^\alpha$ konvergiert nach Satz 1; es liegt **absolute Konvergenz** im Inneren des Quaders Q vor.

(b) Wegen der Konvergenz gilt $\sum\limits_{|\alpha|=0}^{\infty} a_\alpha x^\alpha = \sum\limits_{\alpha_1=0}^{\infty}\left(\sum\limits_{\alpha_2=\cdots=\alpha_n=0}^{\infty} a_\alpha x_2^{\alpha_2}\dots x_n^{\alpha_n}\right) x_1^{\alpha_n} = f(x)$, wobei $x_2,\dots,x_n$ fest sind. Nach 5.20 (Differentiation von Reihen einer Variablen) existiert $\frac{\partial f}{\partial x_1}$ und ist gleich $\frac{\partial f}{\partial x_1}(x) = \sum\limits_{\alpha_1=1}^{\infty}\left(\sum\limits_{\alpha_1\dots\alpha_2=0}^{\infty} a_\alpha x_2^{x_2}\dots x_n^{\alpha_n}\right)\alpha_1 x_1^{\alpha_1-1}$: $\frac{\partial f}{\partial x_1}$ ist wieder stetig, also spielt die Differentiationsreihenfolge keine Rolle. Mittels Induktion behandelt man $f\in C^k(Q),\ k=1,2,\dots,$ bzw. $f\in C^\infty(Q)$.

Die letzte Formel ergibt sich durch Differentiation. □

Korollar. Konvergiert die Potenzreihe $f(x) := \sum\limits_{|\alpha|=0}^{\infty} a_\alpha (x-x_0)^\alpha$ in einem offenen Quader um x_0 , stimmt sie mit der Taylorreihe von f überein.

Satz 4. *Sei $x_0\in\mathbb{R}^n,\ c\in\mathbb{R}^n$ mit $c_j>0, j=1,\dots,n$ und $Q := \{x\in\mathbb{R}^n \mid |x_j-x_{0j}|<c_j,\ j=1,\dots,n\}$. Sei $f\in C^\infty(Q)$ und es gebe $R>0$ mit: $\frac{|D_\alpha f(x)|}{\alpha!}c^\alpha \le R$ gilt für alle $x\in Q$. Dann wird f in Q durch seine Taylorreihe dargestellt, $f(x) = \sum\limits_{|\alpha|=0}^{\infty} \frac{D_\alpha f(x_0)}{\alpha!}(x-x_0)^\alpha$.*

Beweis. O.B.d.A. sei $x_0=0,\ x\in Q$.

$$|f(x) - \sum_{|\alpha|=0}^{k-1}\frac{D_\alpha f(0)}{\alpha!}x^\alpha| = \left|\sum_{|\alpha|=k}\frac{D_\alpha f(\theta x)}{\alpha!}x^\alpha\right| \le \sum_{|x|=k}\frac{|D_\alpha f(\theta x)|}{\alpha!}c^\alpha\left|\frac{x^\alpha}{c^\alpha}\right|$$
$$\le R\sum_{|\alpha|=k} q^\alpha \text{ mit } q=(q_1\dots q_n),\ q_i := \frac{|x_i|}{c_i} < 1\ .$$

Für $k\to\infty$ geht der Fehler (Rest) gegen Null.

□

7.11 Konstanten

Definition. Sei $\Omega\subseteq\mathbb{R}^n$ offen. Ω heißt **zusammenhängend** :⇔ Je zwei Punkt $x_0,x\in\Omega$ lassen sich durch einen Streckenzug in Ω verbinden. $\Omega\subseteq\mathbb{R}^n$ Gebiet :⇔ Ω ist offen und zusammenhängend.

Satz. *Sei $\Omega\subseteq\mathbb{R}^n$ ein Gebiet und $f:\Omega\to\mathbb{R}$ in $C^1(\Omega)$ mit $D_jf=0$ in Ω $(j=1,\dots,n)$. Dann ist f konstant in Ω .*

Beweis. Seien $x_0,x\in\Omega,\ x_0,x_1,x_2,\dots,x_{\ell-1},x_\ell=x$ die Ecken eines x_0 und x verbindenden Streckenzugs in Ω . Auf jede Teilstrecke von x_{j-1} nach x_j $(j=1,\dots,\ell)$ angewandt, ergibt der Mittelwertsatz, dass $f(x_j)=f(x_{j-1})$ ist. Somit ist $f(x)=f(x_0)$ konstant. □

7.12 Definite Matrizen

Wie im Eindimensionalen lassen sich der Mittelwertsatz und die Taylorsche Formel zur Charakterisierung von Maxima und Minima von Funktionen $f : \Omega \subseteq \mathbb{R}^n \to \mathbb{R}$ verwenden. Dazu benötigen wir einige Aussagen aus der linearen Algebra.

Definition. Sei $A = (a_{ij})_{i,j=1}^n$ eine symmetrische Matrix. A heißt **positiv [negativ] definit** $:\Leftrightarrow \forall x \in \mathbb{R}^n - \{0\} \quad \sum\limits_{i,j=1}^n a_{ij}x_ix_j > 0 \ [< 0]$. A heißt **semidefinit**, falls dies ≥ 0 $[\leq 0]$ ist. Man nennt $\Delta_k := \det (a_{ij})_{i,j=1}^k$ die **k-te Abschnittsdeterminante** von A für $1 \leq k \leq n$.

Satz 1. *Sei A eine symmetrische $n \times n$-Matrix. Dann gilt:*

(1) *A ist positiv definit $\Leftrightarrow$ $\Delta_k > 0,\ k = 1, \dots, n$.*

(2) *A ist negativ definit $\Leftrightarrow$ $(-1)^k \Delta_k > 0,\ k = 1, \dots, n$ (d.h. die Δ_k haben alternierende Vorzeichen).*

Satz 2. *Sei A eine symmetrische $n \times n$-Matrix. Dann gilt:*
A ist positiv definit $\Leftrightarrow$ $\exists \alpha > 0\ \forall x \in \mathbb{R}^n \quad \sum\limits_{i,j=1}^n a_{ij}x_ix_j \geq \alpha ||x||_2^2$.

Beweis. "$\Leftarrow$" klar.
"$\Rightarrow$ " Betrachte $f(x) = \sum\limits_{i,j=1}^n a_{ij}x_ix_j$ für $x = (x_j)_{j=1}^n \in \mathbb{R}^n$. Dann ist f stetig auf dem $\mathbb{R}^n$, nimmt also ihr Maximum und Minimum auf der kompakten Menge
$S = \{x \in \mathbb{R}^n \mid ||x||_2 = 1\}$ an. Wähle $x_0 \in S$ mit $f(x_0) = \min\limits_{x \in S} f(x)$. Da $x_0 \neq 0$, muss $\alpha := f(x_0) > 0$ sein, denn A ist positiv definit. Somit gilt $\frac{1}{||x||_2^2} f(x) = f\left(\frac{x}{||x||_2}\right) \geq \alpha$, d.h. $f(x) \geq \alpha ||x||_2^2$.

□

7.13 Extremalprobleme

Definition. Sei $\Omega \subseteq \mathbb{R}^n$ offen. Eine stetige Funktion $f : \Omega \to \mathbb{R},\ f \in C(\Omega)$ besitzt in $x_0 \in \Omega$ ein relatives Maximum [Minimum]
$:\Leftrightarrow\ \exists x_0 \in U \subseteq \Omega$ Umgebung $\forall x \in U, x \neq x_0 \quad f(x_0) \geq f(x)\ [f(x_0) \leq f(x)]$.
Gilt sogar $>$ $[<]$, heißt das Maximum [Minimum] **strikt**. Ein **Extremum** ist ein Maximum oder ein Minimum. **Absolute** Extrema sind analog definiert.

Satz 1. *Sei $\Omega \subseteq \mathbb{R}^n$ eine Umgebung von x_0 und sei $f : \Omega \to \mathbb{R}$ in $C^1(\Omega)$. Dann gilt:*
Hat f in x_0 ein relatives Extremum, folgt $\nabla f(x_0) = 0$.

Beweis. Sei $j \in \{1, \dots, n\}$. Dann hat $g(x_j) := f(x_{0,1}, \dots, x_{0,j-1}, x_j, x_{0,j+1}, \dots, x_{0,n})$ in $x_j = x_{0,j}$ ein relatives Extremum. Also ergibt die Differentialrechnung in einer Variablen, dass $\frac{\partial g}{\partial x_j}(x_0) = 0$ für alle j ist.

□

$\nabla f(x_0) = 0$ ist notwendig, aber nicht hinreichend. Für $f \in C^2(\Omega)$ erhält man folgendes hinreichende Analogon der Theorie einer Variablen:

Satz 2. *Sei* $\Omega \subseteq \mathbb{R}^n$ *eine Umgebung von* x_0*, und sei* $f : \Omega \to \mathbb{R}$ *in* $C^2(\Omega)$ *. Es gelte* $\nabla f(x_0) = 0$ *.*

(1) *Falls die* **Hessesche Matrix** $Hf(x_0) := (D_i D_j f(x_0))_{i,j=1}^n$ *positiv (negativ) definit ist, besitzt* f *in* x_0 *ein striktes relatives Minimum (Maximum).*

(2) *Besitzt* f *in* x_0 *ein relatives Minimum (Maximum), so ist* $Hf(x_0)$ *positiv (negativ) semidefinit.*

Definition. Die Nullstellen x_0 von ∇f heißen **kritische Punkte von** f.

Beweis. In einer kleinen Umgebung $||x-x_0|| = ||h|| < \varepsilon$ gilt nach der Taylorschen Formel $f(x_0+h) = f(x_0) + \nabla f(x_0)h + \frac{1}{2} \sum\limits_{i,j=1}^{n} D_i D_j f(x_0+\theta h) h_i h_j$ mit $0 < \theta < 1$, also mit der Voraussetzung $\nabla f(x_0) = 0$

$$f(x_0+h) - f(x_0) = \frac{1}{2} \sum_{i,j=1}^{n} Hf(x_0+\theta h)_{ij} h_i h_j \ .$$

Nach dem Satz von Schwarz (7.2) ist $Hf(y)$ symmetrisch.

(1) Wenn $Hf(x_0)$ positiv definit ist, gibt es nach 7.12 ein $\alpha > 0$ mit:
$\sum_{i,j=1}^n Hf(x_0)_{ij} h_i h_j \geq \alpha ||h||_2^2$ für alle $h \in \mathbb{R}^n$. Da $f \in C^2(\Omega)$ ist, hängt $Hf(x)$ stetig von x ab, d.h. für alle $1 \leq i,j \leq n$ gilt: $Hf(x)_{ij} \to Hf(x_0)_{ij}$ für $x \to x_0$. Somit gilt in einer kleinen ε-Umgebung von x_0, $||h||_2 < \varepsilon$, dass

$$\sum_{i,j=1}^{n} Hf(x_0+\theta h)_{ij} h_i h_j \geq \frac{\alpha}{2} ||h||_2^2 > 0$$

ist. Es folgt $f(x_0+h) > f(x_0)$: f hat in x_0 ein striktes Minimum. Der Fall des Maximums ist analog.

(2) f habe in x_0 ein relatives Minimum. Wäre $Hf(x_0)$ nicht positiv semidefinit, gäbe es ein $y \in \mathbb{R}^n - \{0\}$ mit $\sum\limits_{i,j=1}^{n} Hf(x_0)_{ij} y_i y_j < 0$. Aus der Stetigkeit von $Hf(\cdot)$ folgt: Für alle $||h||_2 < \varepsilon$ für hinreichend kleines $\varepsilon > 0$ und ein geeignetes $0 < \theta < 1$ ist $\sum\limits_{i,j=1}^{n} Hf(x_0+\theta h)_{ij} y_i y_j < 0$, d.h. für $h = \lambda y$ mit $\lambda \in \mathbb{R}^+$ genügend klein ist $f(x_0+h) - f(x_0) < 0$, d.h. f hätte kein relatives Minimum in x_0 .

□

Im Spezialfall $n = 2$ liefert eine Kombination der Sätze 2 aus 7.12 und 7.13 das folgende Kriterium.

Satz 3. *Sei* $f : \Omega \subseteq \mathbb{R}^2 \to \mathbb{R}$ *in* C^2 *mit* $f_x(x_0,y_0) = f_y(x_0,y_0) = 0$ *.*

a) *Gilt* $f_{xx} f_{yy} - f_{xy}^2 > 0$ *und* $f_{xx} \neq 0$ *in* (x_0,y_0) *, besitzt* f *in* (x_0,y_0) *ein striktes relatives Extremum, und zwar ein Minimum für* $f_{xx} > 0$ *und ein Maximum für* $f_{xx} < 0$ *.*

b) *Gilt* $f_{xx}f_{yy} - f_{xy}^2 < 0$ *in* (x_0, y_0) *, hat* f *in* (x_0, y_0) **kein** *relatives Extremum in* (x_0, y_0) *.*

Beispiele. Nicht immer sind die Kriterien anwendbar, aber doch recht brauchbar.

1. Betrachte $f : \mathbb{R}^2 \to \mathbb{R}, f(x,y) := x^3 + y^3 - 3xy$. Die partiellen Ableitungen $f_x(x,y) = 3x^2 - 3y = 0,\ f_y(x,y) = 3y^2 - 3x = 0$ liefern die kritischen Punkte $(0,0)$ und $(1,1)$. Man hat

$$f_{xx}(x,y) = 6x,\ f_{yy}(x,y) = 6y,\ f_{xy}(x,y) = f_{yx}(x,y) = -3\ .$$

Also ist $(f_{xx}f_{yy} - f_{xy}^2)(0,0) = 3,\ f_{xx}(0,0) = 0$: das Kriterium ist **nicht** anwendbar. Aber in $(0,0)$ liegt **kein** Extremum vor, da

$$f(x,x) = x^2(2x-3) < 0\ ,\ 0 < x < 3/2\ ;\ f(x,-x) = 3x^2 > 0 \text{ für } x \neq 0\ .$$

In $(1,1)$ gilt $(f_{xx}f_{yy} - f_{xy}^2)(1,1) = 27 > 0,\ f_{xx}(1,1) = 6 > 0$, d.h. f hat in $(1,1)$ relatives Minimum, $f(1,1) = -1$. Die Funktion f hat **kein** absolutes Extremum, da $f(x,x) \to \pm\infty$ für $x \to \pm\infty$.

2. $f : \mathbb{R}^2 \to \mathbb{R}, f(x,y) = \sqrt{x^2+y^2}$ hat natürlich ein absolutes Minimum in $(0,0)$, sonst kein relatives Maximum oder Minimum. Aber f ist in $(0,0)$ nicht differenzierbar und das Kriterium nicht anwendbar.

3. $f : D \to \mathbb{R}, f(x,y) = \sin x\ \sin y\ \sin(x+y), D := \{(x,y) \in \mathbb{R}^2 \mid 0 \leq x, y, x+y \leq \pi\}$. $f\mid_{\partial D} = 0$ und $f \geq 0$: f hat ein lokales und globales Minimum auf ∂D und f ist > 0 im Inneren von D . Es gilt für die kritischen Punkte (x,y) von f :

$$\begin{aligned} f_x(x,y) &= \cos x \sin y \sin(x+y) + \sin x \sin y \cos(x+y) = 0\ , \\ f_y(x,y) &= \sin x \cos y \sin(x+y) + \sin x \sin y \cos(x+y) = 0\ . \end{aligned}$$

Also: $\frac{\cos x}{\cos y} = \frac{\sin x}{\sin y}$, d.h. $\tan y = \tan x$ durch Division (aus $\sin x, \sin y \neq 0$ folgt $\cos x, \cos y, \cos(x+y) \neq 0$). Damit folgt $y = x$, also $\cos x \sin 2x + \sin x \cos 2x = 0$, d.h. $\sin 3x = 0,\ x = \pi/3$. Also ist nur $(\pi/3, \pi/3)$ eine kritische Stelle im Inneren von D . Die Rechnung zeigt $f_{xx}(\pi/3,\pi/3) = -\sqrt{3}$, $(f_{xx}f_{yy} - f_{xy}^2)(\pi/3,\pi/3) = 9/4 > 0$. Also hat f in $(\pi/3,\pi/3)$ ein Maximum, das ein absolutes Maximum ist. Es gilt $f(\pi/3,\pi/3) = 3\sqrt{3}/8$.

4. $f(x,y) := (y - x^2)(y - 3x^2)$ hat **kein** relatives Extremum in (0,0); f ist positiv in Punkten der Form $(0,b)$, negativ in Punkten $(a, 2a^2)$ für $a, b \in \mathbb{R}$. Wird f jedoch auf $x = 0,\ y = 0$ oder $y = mx,\ m \in \mathbb{R}$ eingeschränkt, hat f ein striktes relatives Minimum in 0. Die Funktion g, gegeben durch $g(x) = m^2x^2 - 4mx^3 + 3x^4$, erfüllt nämlich $g'(0) = 0,\ g''(0) = 2m^2 > 0$.

5. Minimiere $f(a,b) = \sum_{k=1}^{n}(a + bx_k - y_k)^2$: als Lösung ergibt sich die Ausgleichsgerade $y = a + bx$ für Paare $\{(x_k, y_k),\ k = 1, \ldots, n\}$.

7.14 Der Banachsche Fixpunktsatz

Satz (Banach). *Sei D ein vollständiger metrischer Raum und sei $f : D \to D$* **kontrahierend**, *d.h. $\exists k < 1\ \forall x, y \in D \quad d(f(x), f(y)) \leq k\ d(x, y)$. Dann besitzt f genau einen Fixpunkt $\hat{x}$: $\exists^1 \hat{x} \in D$ mit $f(\hat{x}) = \hat{x}$. Der Fixpunkt $\hat{x}$ ist berechenbar über: Sei $x_1 \in D$ beliebig, $x_{n+1} := f(x_n)$. Dann konvergiert $x_n \to \hat{x}$ mit $d(x_n, \hat{x}) \leq \frac{k}{1-k}\ d(x_n, x_{n-1})$. Der Satz gilt speziell für abgeschlossene Teilmengen $D \subseteq X$ von Banachräumen X.*

Beweis. Sei $x_1 \in D$ und $x_{n+1} := f(x_n)$ induktiv definiert.

Behauptung. $(x_n)_{n\in\mathbb{N}}$ ist Cauchyfolge in D . Es ist für $n \in \mathbb{N}, n \geq 2$,

$$d(x_n, x_{n-1}) = d(f(x_{n-1}), f(x_{n-2})) \leq k\ d(x_{n-1}, x_{n-2}) \leq \ldots k^{n-2} d(x_2, x_1)\ .$$

Also gilt für $n > m \geq 1$

$$d(x_n, x_m) \leq \sum_{j=m+1}^{n} d(x_j, x_{j-1}) \leq \left(\sum_{j=m+1}^{n} k^{j-2} \right) d(x_2, x_1)\ .$$

Die geometrische Reihe $\sum\limits_{j\in\mathbb{N}_0} k^j$ konvergiert, da $k < 1$ ist. Somit ist $(x_n)_{n\in\mathbb{N}}$ Cauchyfolge. Also folgt: $x_n \to \hat{x} \in D$ ist konvergent in D . Der Grenzwert $\hat{x}$ ist Fixpunkt von f , denn

$$\hat{x} = \lim_n x_n = \lim_n f(x_{n-1}) = \lim_n f(x_n) = f(\hat{x})\ ,$$

da f stetig, weil kontrahierend ist. Der Fixpunkt ist eindeutig: ist auch $\hat{x}' \in D$ Fixpunkt, gilt $d(\hat{x}, \hat{x}') = d(f(\hat{x}), f(\hat{x}')) \leq k\ d(\hat{x}, \hat{x}')$. Mit $k < 1$ folgt $d(\hat{x}, \hat{x}') = 0$, also $\hat{x} = \hat{x}'$.
Als Fehlerabschätzung ergibt sich: $d(x_m, x_n) \leq \sum\limits_{j=n+1}^{m} d(x_j, x_{j-1}) \leq \sum\limits_{j=n+1}^{m} k^{j-n} d(x_n, x_{n-1})$.
Für $m \to \infty$ also $d(\hat{x}, x_n) \leq c\ d(x_n, x_{n-1})$ mit $c = \sum\limits_{j=n+1}^{\infty} k^{j-n} = \sum\limits_{i=1}^{\infty} k^i = \frac{k}{1-k}$.

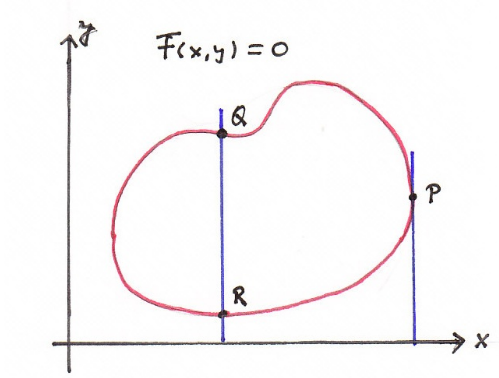
Bemerkung. Der Fixpunktsatz wird häufig zur Lösung von Nullstellenproblemen $g(x) = 0$ angewandt, indem man $f(x) = x - g(x)$ betrachtet: Nullstellen in g sind dann gerade die Fixpunkte von f .

7.15 Implizite Funktionen

In Anwendungen kommen häufig Konstanzlinien von Funktionen $F : \mathbb{R}^2 \to \mathbb{R}$ vor, z.B. Höhenlinien (also Linien konstanter Höhe c) auf Karten, Isobaren (d.h. Linien konstanten Druckes p) auf Wetterkarten, Isothermen (das sind Linien konstanter Temperatur T), d.h. man betrachtet für ein festes $c \in \mathbb{R}$

$$\Gamma_c = \{(x, y) \in \mathbb{R}^2 \mid F(x, y) = c\}\ .$$

Sei o.B.d.A. $c = 0$ (Subtraktion von c) und $\Gamma := \Gamma_0$.



Beispiel. $F(x,y) = x^2 + y^2 - 1 = 0$: Dann ist Γ der Kreis um $(0,0)$ mit Radius 1. In der "Nähe" eines gegebenen Punktes $(x_0, y_0) \in \Gamma$ ist Γ "häufig" der Graph einer **Funktion**, d.h. $\exists \varepsilon, \delta > 0$ und $f : U_\delta(x_0) \to U_\varepsilon(y_0)$ mit $F(x, f(x)) = 0$ für alle $x \in U_\delta(x_0)$: $\begin{pmatrix} x \\ f(x) \end{pmatrix} \in \Gamma$. Implizit wird durch $F(x,y) = 0$ eine Funktion $y = f(x)$ definiert. Man sagt, dass man $F(x,y) = 0$ nach y auflöst. Probleme, die auftreten können, sind:

1) $\Gamma = \emptyset$. Man muss also $\Gamma(x_0, y_0) = 0$ voraussetzen: $(x_0, y_0) \in \Gamma$.

2) Falls $\varepsilon > 0$ zu groß ist, könnte $F(x_0, y) = 0$ zwei Lösungen y_1, y_2 haben (im Bild R, Q), also $f(x_0) = y$ nicht eindeutig (als Funktion!) definiert sein. Gewünscht ist, dass f wohldefiniert und stetig ist, wenn F stetig ist.

3) In Punkten $P = \begin{pmatrix} x_0 \\ y_0 \end{pmatrix} \in \Gamma$ des Graphen, in denen die Tangente vertikal ist, kommen bei noch so kleiner δ- und ε-Wahl stets zu $x \in U_\delta(x_0)$, $x \neq x_0$, zwei Punkte $f_1(x)$ und $f_2(x)$ in $U_\varepsilon(y_0)$ für $F(x, f(x)) = 0$ in Frage. In diesem Fall ist $\frac{\partial F}{\partial y}(x_0, y_0) = 0$ (im obigen Fall des Kreises ist $\frac{\partial F}{\partial y} = 2y$, $\frac{\partial F}{\partial y}(1,0) = 0$, $(1,0) \in \Gamma$), dann ist die Definition von f zumindest problematisch oder unmöglich.

Man wird also $F(x_0, y_0) = 0$ und $\frac{\partial F}{\partial y}(x_0, y_0) \neq 0$ voraussetzen und nur eine lokale Lösbarkeit von $F(x,y) = 0$ durch $y = f(x)$ in einer kleinen Umgebung von (x_0, y_0) erwarten können. Diese **implizite Definition von f** besagt **nicht**, dass man dies rechnerisch-formelmäßig tun kann!

7.16 Der Hauptsatz über implizite Funktionen

Wir formulieren das Problem allgemeiner für Funktionen $F : \mathbb{R}^{n+m} \to \mathbb{R}^m$. Dazu seien einige Bezeichnungen eingeführt. Für $x = {}^t(x_1, \ldots, x_n) \in \mathbb{R}^n$, $y = {}^t(y_1, \ldots, y_m) \in \mathbb{R}^m$ sei ${}^t(x,y) = {}^t(x_1, \ldots, x_n, y_1, \ldots, y_m) \in \mathbb{R}^{n+m}$. Da F $\mathbb{R}^m$-wertig ist, hat F die Form $F = {}^t(F_1, \ldots, F_m)$ mit Koordinatenfunktionen F_i . Man setzt

$$\frac{\partial F}{\partial x} := \begin{pmatrix} \frac{\partial F_1}{\partial x_1} & \cdots & \frac{\partial F_1}{\partial x_n} \\ \vdots & & \\ \frac{\partial F_m}{\partial x_1} & \cdots & \frac{\partial F_m}{\partial x_n} \end{pmatrix}_{m \times n} , \quad \frac{\partial F}{\partial y} := \begin{pmatrix} \frac{\partial F_1}{\partial y_1} & \cdots & \frac{\partial F_1}{\partial y_m} \\ \vdots & & \\ \frac{\partial f_m}{\partial y_1} & \cdots & \frac{\partial F_m}{\partial y_m} \end{pmatrix}_{m \times m} .$$

Also: $\frac{\partial F}{\partial x}$ ist die ("partielle") totale Ableitung von $x \mapsto F(x,y)$ bei festem y und $\frac{\partial F}{\partial y}$ ist die ("partielle") totale Ableitung von $y \mapsto F(x,y)$ bei festem x . Die Gleichung $F(x,y)=0$ **auf** $G \subseteq \mathbb{R}^n$ **nach** y **aufzulösen**, soll heißen:
Finde $f: G \subseteq \mathbb{R}^n \to \mathbb{R}^m$ mit: Für alle $x \in G$ ist $F(x,f(x)) = 0$, d.h. explizit gilt mit $f = {}^t(f_1, \ldots, f_m)$

$$\left\{ \begin{array}{c} F_1(x_1 \ldots x_n, f_1(x_1 \ldots x_n), \ldots f_m(x_1 \ldots x_n)) = 0 \\ \vdots \\ F_m(x_1, \ldots x_n, f_1(x_1 \ldots x_n), \ldots f_m(x_1, \ldots, x_n) = 0 \end{array} \right\} .$$

Zu finden sind also die Funktionen $f_1, \ldots, f_m : G \to \mathbb{R}$.

Hauptsatz über implizite Funktionen: *Seien $G \subseteq \mathbb{R}^n$, $H \subseteq \mathbb{R}^m$ offen und sei $F: G \times H \to \mathbb{R}^m$ stetig differenzierbar. Seien $x_0 \in G$, $y_0 \in H$ gegeben mit:*

$$F(x_0,y_0) = 0 \ , \ \frac{\partial F}{\partial y}(x_0,y_0) \ \textit{ist eine invertierbare } (m \times m) - \textit{Matrix} \ .$$

Dann gibt es $\delta > 0$ und $\varepsilon > 0$ mit $U := U_\delta(x_0) \subseteq G$, $V := U_\varepsilon(y_0) \subseteq H$ und eine Funktion $f: U \to V$, so dass für alle $x \in U$

$$f(x_0) = y_0 \quad \textit{und} \quad F(x,f(x)) = 0$$

ist. Für jedes feste $x \in U$ ist $f(x)$ dabei die einzige in V liegende Lösung von $F(x,y) = 0$. Die Funktion f liegt in $C^1(U_1, V)$ für eine geeignet kleine δ_1-Umgebung U_1 von x_0 , $0 < \delta_1 \leq \delta$ und man hat

$$f'(x) = -\left(\frac{\partial F}{\partial y}(x,f(x))\right)^{-1} \circ \frac{\partial F}{\partial x}(x,f(x)), \ x \in U_1 \ .$$

Beweis.

i) Sei $D := \frac{\partial F}{\partial y}(x_0,y_0)$. Dann ist $D^{-1}\frac{\partial F}{\partial y}(x_0,y_0) = I$ die Identität auf dem $\mathbb{R}^m$. Da $F(x_0,y_0) = 0$, $\frac{\partial F}{\partial y}$ stetig in (x_0,y_0) ist und D^{-1} stetig und linear ist, gibt es $\delta > 0$ und $\varepsilon > 0$ sowie eine δ-Umgebung $U \subset G$ von x_0 und eine ε-Umgebung $V \subseteq H$ von y_0 mit:

$$\begin{aligned} ||I - D^{-1}\frac{\partial F}{\partial y}(x,y)|| &\leq \frac{1}{2} \ \ (x \in U, y \in V) \qquad (1) \\ ||D^{-1}F(x,y_0)|| &\leq \varepsilon/4 \ \ (x \in U), \end{aligned}$$

Setze
$X := C_b(U,\mathbb{R}^m) := (\{f : U \subseteq \mathbb{R}^n \to \mathbb{R}^m \mid f \text{ stetig und beschränkt }\}, \ || \cdot ||_\infty)$. Da $\mathbb{R}^m$ vollständig ist, ist X ein Banachraum unter der Norm $||f||_\infty = \sup\limits_{x \in U} ||f(x)||_2$.
Sei ferner

$$M := \{g \in X \mid g(x_0) = y_0, \ ||g(x) - y_0|| \leq \varepsilon/2 \text{ für alle } x \in U\} \subseteq X := C_b(U, \mathbb{R}^n) \ .$$

Dann ist M eine nicht-leere, abgeschlossene Teilmenge von X; M ist nicht-leer, da $g_0(x) := y_0$ zu M gehört. Wegen $||g(x)|| \leq ||g(x) - y_0|| + ||y_0|| \leq \varepsilon/2 + ||y_0||$ ist g beschränkt.

ii) Die Werte von $g \in M$ liegen also in $V = U_\varepsilon(y_0)$. Definiere

$$A : M \subseteq X \to X \;\text{ durch }\; (Ag)(x) := g(x) - D^{-1}F(x, g(x)) \;.$$

Dann ist $Ag(x_0) = y_0$ und Ag ist stetig. Zunächst gilt wegen (1) für alle $x \in U$

$$||Ag_0(x) - y_0|| = ||D^{-1}F(x, y_0)|| \leq \varepsilon/4 \;. \tag{2}$$

Für **festes** $x \in U$ definiere $\phi : V \to \mathbb{R}^m$ durch $\phi(y) := y - D^{-1}F(x, y)$. Dann ist

$$\phi'(y) = I - D^{-1}\,\frac{\partial F}{\partial y}(x, y) \;.$$

Nach dem Korollar zum Mittelwertsatz (7.8) gilt für alle $y, z \in V$

$$\begin{aligned} ||\phi(y) - \phi(z)|| &\leq \sup_{\tilde{y} \in V} ||\phi'(\tilde{y})||\;||y - z|| \\ &= \sup_{\tilde{y} \in V} ||I - D^{-1}\frac{\partial F}{\partial y}(x, y)||\;||y - z|| \leq \frac{1}{2}||y - z|| \;. \end{aligned}$$

Speziell ergibt sich für alle $g_1, g_2 \in M$

$$||\phi(g_1(x)) - \phi(g_2(x))|| \leq \frac{1}{2}||g_1(x) - g_2(x)|| \;. \tag{3}$$

Aber $\phi(g_j(x)) = g_j(x) - D^{-1}F(x, g_j(x)) = (Ag_j)(x)$, d.h.

$$||Ag_1 - Ag_2||_X \leq \frac{1}{2}||g_1 - g_2||_X \;,\; g_1, g_2 \in M \;. \tag{4}$$

Speziell folgt aus (3) für $g \in M, x \in U$ mit der konstanten Function g_0 aus i)

$$\begin{aligned} ||(Ag)(x) - y_0|| &\leq ||(Ag)(x) - (Ag_0)(x)|| + ||(Ag_0)(x) - y_0|| \\ &\leq \frac{1}{2}||g(x) - y_0|| + \frac{1}{4}\varepsilon \leq \frac{1}{2}\frac{\varepsilon}{2} + \frac{\varepsilon}{4} = \varepsilon/2 \;. \end{aligned}$$

Somit ist $Ag \in M$, d.h. $A : M \to M$ ist eine kontrahierende Selbstabbildung von M , vgl. (4). Aus dem Banachschen Fixpunktsatz (7.14) folgt, dass A einen Fixpunkt $f \in M$ besitzt, $Af = f$. Die Abbildung $f : U \to V$ ist stetig und erfüllt $f(x_0) = y_0$ sowie $D^{-1}F(x, f(x)) = 0$. Durch Anwendung von D ergibt sich $F(x, f(x)) = 0$.

Eindeutigkeit: Bei gegebenen $x \in U$ sei $y \in V$ Lösung von $F(x, y) = 0$. Dann ist $y = f(x)$, da $||y - f(x)|| = ||\phi(y) - \phi(f(x))|| \leq \frac{1}{2}||y - f(x)||$.

iii) Wir zeigen jetzt die Differenzierbarkeit von f in x_0 . Sei o.B.d.A. $x_0 = y_0 = 0$. Wir wählen als Norm auf dem $\mathbb{R}^{n+m}$: $||z|| := ||x|| + ||y||$ für $z = {}^t(x, y)$. Sei $D_1 := \frac{\partial F}{\partial x}(0, 0)$ und $D_2 := \frac{\partial F}{\partial y}(0, 0)$. Da F in (0,0) differenzierbar ist, gilt für z mit kleiner Norm

$$F(z) = F'(0)z + r(z) \;,\; \text{wobei } \frac{r(z)}{||z||} \to 0 \;(z \to 0) \;,$$

d.h. $F(x,y) = D_1x + D_2y + r(x,y)$, $\frac{r(x,y)}{||x||+||y||} \overset{(*)}{\to} 0$ $(x,y \to 0)$. Da für $x \in U$ $F(x,f(x)) = 0$ ist, folgt

$$0 = D_1x + D_2f(x) + r(x,f(x)) \ .$$

Also gilt

$$f(x) = -D_2^{-1}D_1x - D_2^{-1}r(x,f(x)) \ . \qquad (5)$$

Falls also $\lim\limits_{x\to 0} D_2^{-1}\frac{r(x,f(x))}{||x||} = 0$ ist, folgt, dass f in 0 differenzierbar ist mit $f'(0) = -D_2^{-1}D_1$. Dazu reicht es, $\lim\limits_{x\to 0} \frac{r(x,f(x))}{||x||} = 0$ zu beweisen. Wegen $(*)$ gibt es $0 < \delta_1 \le \delta$ und $0 < \varepsilon_1 \le \varepsilon$, so dass für alle $x \in \mathbb{R}^n, y \in \mathbb{R}^m$ mit $||x|| < \delta_1$ und $||y|| < \varepsilon_1$ gilt

$$||r(x,y)|| \le \frac{1}{2||D_2^{-1}||}(||x| + ||y||) \ .$$

Da f stetig in 0 ist, existiert $\delta_1' \le \delta_1$, so dass für alle $x \in \mathbb{R}^n$ mit $||x|| \le \delta_1'$ gilt $||f(x)|| \le \varepsilon_1$. Dafür gilt dann wegen (5)

$$||f(x)|| \ \le \ ||D_2^{-1}D_1||\ ||x|| + ||D_2^{-1}||\frac{||x|| + ||f(x)||}{2||D_2^{-1}||} = \frac{\gamma}{2}||x|| + \frac{1}{2}||f(x)|| \ ,$$

$$||f(x)|| \le \gamma||x||; \ ||x|| < \delta_2; \gamma := 2||D_2^{-1}D_1|| + 1 \ . \qquad (6)$$

Also $0 \le \frac{||r(x,f(x))||}{||x||} = (1+\gamma)\ \frac{||r(x,f(x))||}{||x||+\gamma||x||} \le (1+\gamma)\ \frac{||r(x,f(x))||}{||x||+||f(x)||}$. Für $x \to 0$ gilt auch $f(x) \to 0$, somit $\frac{||r(x,f(x))||}{||x||+||f(x)||} \to 0$. Damit ist die Differenzierbarkeit von f in x_0 bewiesen.

iv) Da $\frac{\partial F}{\partial y}$ stetig in (x_0,y_0) ist, gibt es einen δ_2-Umgebung $U_2 \subseteq U$ von x_0 und eine ε_1-Umgebung $V_1 \subseteq V$ von y_0 mit:

$$||\frac{\partial F}{\partial y}(x,y) - \frac{\partial F}{\partial y}(x_0,y_0)|| < \frac{1}{||\frac{\partial F}{\partial y}(x_0,y_0)^{-1}||} \ ; \ x \in U_2, \ y \in V_1 \ .$$

Mit der Neumannschen Reihe folgt, dass $\frac{\partial F}{\partial y}(x,y)$ invertierbar ist.

Man zeigt nämlich mit der **Neumannschen Reihe**: $||A - B|| < \frac{1}{||A^{-1}||}$ impliziert $B^{-1} = \sum\limits_{n=0}^{\infty} A^{-1}(I - BA^{-1})^n$. Da f stetig ist, gibt es eine δ_3-Umgebung $U_1 \subseteq U_2$ von x_0 mit
$f(U_1) \subseteq V_1$. Nach i) – iii), angewandt auf $(x,f(x))$ statt auf (x_0,y_0) ist also f in $x \in U_1$ differenzierbar mit

$$f'(x) = -\frac{\partial F}{\partial y}(x,f(x))^{-1} \cdot \frac{\partial F}{\partial x}(x,f(x)) \ ,$$

und $f'(x)$ hängt stetig von x ab, da $\frac{\partial F}{\partial y}(x,f(x))$ dies tut. □

Bemerkung: Es kann also $f'(x_0)$ berechnet werden, ohne dass f bekannt ist! Nachdem gezeigt ist, dass $f'(x_0)$ existiert, merkt man sich die Differenziationsformel als Anwendung der Kettenregel auf $F(x, f(x)) = 0$:

$$\frac{\partial F}{\partial x}(x_0, y_0) + \frac{\partial F}{\partial y}(x_0, y_0) f'(x_0) = 0 \ .$$

Beispiel. $F(x, y) = e^{2x-y} + 3x - 2y - 1 = 0$ erfüllt $F(0, 0) = 0,\ F_y(0, 0) \neq 0$, also ist $F(x, y) = 0$ nahe (0,0) lokal nach $y = f(x)$ auflösbar.

7.17 Der Umkehrsatz

Wann besitzt $f : G \subseteq \mathbb{R}^n \to \mathbb{R}^n$ eine Inverse f^{-1} ?

Umkehrsatz: Sei $G \subseteq \mathbb{R}^n$ offen und $f : G \to \mathbb{R}^n$ stetig differenzierbar. In $x_0 \in G$ sei $f'(x_0)$ invertierbar. Dann gibt es eine offene Umgebung $W \subseteq G$ von x_0 und eine ε-Umgebung V von $y_0 := f(x_0)$, so dass $f : W \to V$ bijektiv ist. Die Umkehrung $g = f^{-1} : V \to W$ ist stetig differenzierbar mit $g'(y_0) = f'(x_0)^{-1}$.

Beweis. Wir wenden 7.16 auf die Funktion $F : G \times \mathbb{R}^n \to \mathbb{R}^n, F(x, y) := f(x) - y$ an, allerdings mit vertauschten Rollen von x und y und $m = n$. Man hat $F \in C^1$, $F(x_0, y_0) = 0$ und $\frac{\partial F}{\partial x}(x_0, y_0) = f'(x_0)$ ist invertierbar nach Voraussetzung.
Nach 7.16 ist $F(x, y) = 0$ lokal nach x auflösbar, d.h. es gibt eine ε-Umgebung V von y_0 und eine δ-Umgebung $U \subseteq G$ von x_0 sowie genau eine stetige Funktion $g : V \to U$ mit

$$g(y_0) = x_0 \ ,\ F(g(y), y) = 0,\ \text{d.h. } f(g(y)) = y,\ \text{für alle } y \in V \ .$$

O.B.d.A. sei V so klein, dass g gemäß 7.16 in V stetig differenzierbar ist. Wegen der Stetigkeit von f ist $W := U \cap f^{-1}(V) = \{x \in U \mid f(x) \in V\} \subseteq U$ offen mit $f(W) = V$. Aus $y \in V,\ y = f(g(y))$ folgt sogar $f(W) = V$, da $g(V) \subseteq U \cap f^{-1}(V) = W$. f ist injektiv, da $f(x_1) = f(x_2) =: y$ für $x_1, x_2 \in W$ impliziert: $F(x_1, y) = F(x_2, y)$, woraus $x_1 = x_2$ folgt. Also ist $f : W \to V$ bijektiv und g ist stetig differenzierbar als $(f|_W)^{-1}$. Aus der Kettenregel und $(f \circ g)(y) = y$ ergibt sich

$$(f \circ g)'(y) = f'(g(y))g'(y) = I \ ,\ \text{also } g'(y) = f'(x)^{-1} \ .$$

□

Beispiele.

(1) $f : \mathbb{R}^2 \to \mathbb{R}^2,\ f_1(x, y) = x^2 - y^2,\ f_2(x, y) = 2xy,\ \det f'(x, y) = 4(x^2 + y^2) \neq 0$ für $(x, y) \neq (0, 0)$. Also ist f außerhalb des Nullpunktes (0,0) lokal umkehrbar. Aber f ist nicht **global** umkehrbar, da f **nicht** injektiv ist, denn $f(x, y) = f(-x, -y)$.

(2) $f : \mathbb{R}^2 \to \mathbb{R}^2,\ f_1(x, y) = e^x \cos y,\ f_2(x, y) = e^x \sin y$,

$$f'(x, y) = \begin{pmatrix} e^x \cos y & -e^x \sin y \\ e^x \sin y & e^x \cos y \end{pmatrix} \ .$$

Also gilt $\det f'(x,y) = e^{2x} \neq 0$, und f ist überall **lokal** umkehrbar. Aber: f ist **nicht** global umkehrbar, da f nicht injektiv ist: Man hat ja

$$f(x,y) = f(x, y + 2k\pi) \ , \ k \in \mathbb{Z} \ .$$

7.18 Extrema unter Nebenbedingungen

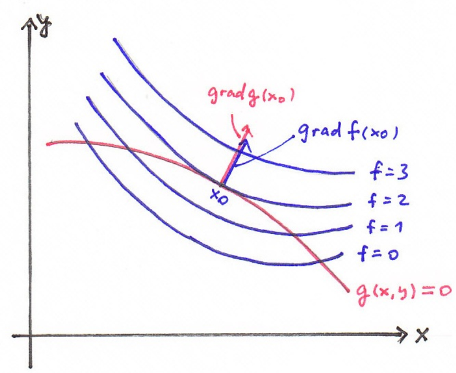
Aufgabe: Bestimme die Extrema von $f : \mathbb{R}^n \to \mathbb{R}$ unter Nebenbedingungen $g_j(x) = 0$, $g_j : \mathbb{R}^n \to \mathbb{R}, \ j = 1, \ldots, m$. **Z.B.** Minimaler Abstand zu Null auf einer "Fläche" $g(x) = 0$.

Satz (Lagrangesche Multiplikationsregel)**.** *Sei $G \subseteq \mathbb{R}^n$ offen und seien $f : G \to \mathbb{R}$ und $g : G \to \mathbb{R}^m$ mit $m < n$ stetig differenzierbar. Die Matrix $g'(x_0) \in L(\mathbb{R}^n, \mathbb{R}^m)$ besitze in $x_0 \in G$ Höchstrang m, $g(x_0) = 0$. f besitze in x_0 ein lokales Extremum unter den Nebenbedingungen $g(x) = 0$. Dann gibt es $\lambda_1, \ldots, \lambda_m \in \mathbb{R}$, so dass gilt:*

$$\frac{\partial f(x_0)}{\partial x_j} + \sum_{i=1}^{m} \lambda_i \frac{\partial g_i(x_0)}{\partial x_j} = 0 \ , \ j = 1, \ldots, n \ . \tag{1}$$

Ferner ist $g_i(x_0) = 0, \ i = 1, \ldots, m$ *.*

Bemerkung: Der Satz liefert nur eine notwendige Bedingung für ein Extremum. Ob ein Extremum in x_0 vorliegt, muss separat geprüft werden. In (1) hat man $n+m$ Gleichungen für $n+m$ Unbekannte $x_{0,1}, \ldots, x_{0,n}, \ \lambda_1, \ldots, \lambda_m$, mit $x_0 = (x_{0,1}, \cdots, x_{0,n})$.



Beweis. O.B.d.A. sei

$$\det \begin{pmatrix} \frac{\partial g_1(x_0)}{\partial x_1} & \cdots & \frac{\partial g_1(x_0)}{\partial x_m} \\ \vdots & & \vdots \\ \frac{\partial g_m(x_0)}{\partial x_1} & \cdots & \frac{\partial g_m(x_0)}{\partial x_m} \end{pmatrix} \neq 0 \ .$$

Nach dem Hauptsatz für implizite Funktionen lässt sich $g(x) = 0$ lokal nach $x_1, \dots, x_m$ auflösen, d.h.

$$x_k = h_k(x_{m+1}, \dots, x_n)\ ,\ k = 1, \dots, m\ . \tag{2}$$

Mit $x = {}^t(x_1, \dots, x_n)$, $y = {}^t(x_1, \dots, x_m)$, $z = {}^t(x_{m+1}, \dots, x_n)$ ist $x = {}^t(y, z)$. (2) bedeutet für $h : U \subseteq \mathbb{R}^{n-m} \to \mathbb{R}^m$, definiert auf einer geeigneten Umgebung U von z_0 , mit $x_0 = {}^t(y_0, z_0)$,

$$g(h(z), z) = 0,\ z \in U \tag{2'}$$

und

$$h(z_0) = y_0,\ \det \frac{\partial g}{\partial y}(x_0) \neq 0\ . \tag{3}$$

Da $(z \mapsto f(h(z), z))$ als Funktion von $z = {}^t(x_{m+1}, \dots, x_n)$ ein relatives Extremum in z_0 besitzt, ist die Ableitung in z_0 gleich 0, d.h.

$$\frac{\partial f}{\partial y}(x_0)\ \frac{\partial h}{\partial z}(z_0) + \frac{\partial f}{\partial z}(x_0) = 0\ . \tag{4}$$

Das lineare Gleichungssystem

$${}^t(\frac{\partial f}{\partial y}(x_0)) + {}^t(\frac{\partial g}{\partial y}(x_0))\lambda = 0\ ,\ \lambda = \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_m \end{pmatrix} \tag{5}$$

ist wegen (3) eindeutig lösbar. Dies sind die ersten m behaupteten Gleichungen.
Die Differentiation von (2') ergibt

$$\begin{aligned} &\frac{\partial g}{\partial y}(x_0)\ \frac{\partial h}{\partial z}(z_0) + \frac{\partial g}{\partial z}(x_0) = 0 \\ &\frac{\partial f}{\partial z}(x_0) \overset{(4)}{=} -\frac{\partial f}{\partial y}(x_0)\frac{\partial h}{\partial z}(z_0) \overset{(5)}{=} {}^t\lambda\frac{\partial g}{\partial y}(x_0)\ \frac{\partial h}{\partial z}(z_0) \overset{(6)}{=} -\ {}^t\lambda\frac{\partial g}{\partial z}(x_0)\ , \end{aligned} \tag{6}$$

Folglich ist ${}^t\big(\frac{\partial f}{\partial z}(x_0)\big) + {}^t\big(\frac{\partial g}{\partial z}(x_0)\big)\lambda = 0$: dies sind die letzten $(n-m)$ der behaupteten Gleichungen.

□

Bemerkung. Die Gleichungen (1) erhält man, indem man formal die Extrema von $L(x, \lambda) = f(x) + \lambda_1 g_1(x) + \dots + \lambda_m g_m(x)$ zu bestimmen versucht. $(\nabla L = 0)$, als Funktion der $(n+m)$ Unbekannten (x, λ) . L heißt auch **Lagrange-Funktion** und die $(\lambda_1, \dots, \lambda_m)$ heißen **Lagrange-Multiplikatoren**.

Beispiele.

(1) Bestimme die Extrema von $f(x, y, z) := 5x + y - 3z$ auf dem Schnitt der Ebene $x + y + z = 0$ mit der Kugeloberfläche $x^2 + y^2 + z^2 = 1$: Bilde

$$L(x, y, z, \lambda, \mu) = (5x + y - 3z) + \lambda(x + y + z) + \mu(x^2 + y^2 + z^2 - 1)$$

Die Differentiation von L ergibt:

$$\left\{\begin{array}{lc} 5+\lambda+2\mu x=0 & \\ & x+y+z=0 \\ 1+\lambda+2\mu y=0 \quad , & \\ & x^2+y^2+z^2=1 \\ -3+\lambda+2\mu z=0 & \end{array}\right\} \qquad (*)$$

Die Addition der ersten drei Gleichungen ergibt: $3+3\lambda=-2\mu(x+y+z)=0$, $\lambda=-1$. Aus der ersten Gleichung folgt: $2+\mu x=0,\ \mu y=0\ :\ \mu\neq 0,\ y=0$.

Die letzte Gleichung impliziert:

$$x=-z,\ 2x^2=1,\ x=\pm\frac{1}{\sqrt{2}}\ .$$

Die Punkte $\left(\frac{1}{\sqrt{2}},0,-\frac{1}{\sqrt{2}}\right)$, $\left(-\frac{1}{\sqrt{2}},0,\frac{1}{\sqrt{2}}\right)$ erfüllen $(*)$ mit $\lambda=-1$ und $\mu=\mp 2\sqrt{2}$. Man hat $f\left(\frac{1}{\sqrt{2}},0,-\frac{1}{\sqrt{2}}\right)=4\sqrt{2},\ f\left(-\frac{1}{\sqrt{2}},0,\frac{1}{\sqrt{2}}\right)=-4\sqrt{2}$: Hier liegen ein Maximum und ein Minimum vor.

(2) **Übung:** Punkt auf Paraboloid $z=x^2/4+y^2/9$ am nächsten an (1,0,0) : $(a,0,a^2/4)$ mit $a^3+8a-8=0,\ a\simeq 0,9068$.

(3) **Übung:** (Hölder-Ungleichung): $f(x)=\sum\limits_{i=1}^{n}a_ix_i,\ \sum\limits_{i=1}^{n}x_i^q=1$ mit $1<q<\infty$, $x_i\geq 0,\ i=1,\dots,n$.

(4) **Übung**: $f(x)=x_1\dots x_n$ in $||x||_2\leq 1$.

(5) **Übung**: Hadamardsche Determinantenabschätzung (Blatter III, S. 39).